

End of the European dream?

Europe's faltering vision of its future has been brought about by social and historical forces for which it had not bargained; the need now is for a clear statement of its goals.

JUST a year ago, in Europe and elsewhere, it was generally believed that once the Maastricht Treaty had been ratified by the twelve members of the European Communities (EC), there would be nothing to stop the further progress towards "ever closer union" specified by the Treaty of Rome more than four decades ago. So, with Denmark aligned behind the treaty after last month's second referendum, and with the British government likely now to follow suit, there would seem to be nothing to prevent the movement for European cohesion (on whatever constitutional basis) from being carried forward by the unspent momentum of the past four decades. Yet that is not how it seems in Europe.

What has gone wrong has little directly to do with the research community, which is not to say that the consequences will be irrelevant to its long-term interests. Three groups of structural problems are separately conspicuous in the present scene. While European (and other) governments naturally seek political solutions, it may be that social scientists and even historians have more constructive things to say about the present ennui.

The most obvious present problem is economic. That is ironic because this is supposed to be the year in which the economic benefits of Europe's single market become apparent. In due course, people within the EC will no doubt benefit enormously from the genuine free trade the single market brings. The snag is that governments have not prepared their people, or even themselves, for the changes the single market will bring. It is not merely a licence for, but a spur to, the amalgamation of enterprises or their relocation, meaning social disruption and even increased unemployment somewhere or other. But this comes about when there are already more than 10 million people out of work in the EC and when the costs of German reunification within Germany have driven up interest rates and unemployment in the countries still belonging to the European Monetary Union — and have made those driven out despair of ever rejoining. Worse, it has begun to dawn on people that Europe has made itself uncompetitive in two important ways — first, employment costs (wages and social transfers) in prosperous member states such as Germany are too high to be competitive with those in the Pacific and even the United States and, second, that Europe can no longer afford the \$50 billion a year it spends on the Common Agricultural Policy, but could hardly afford the cost in unemployment of doing without it.

The second set of problems is even more daunting, and springs from the way in which people are on the move

throughout Eastern Europe and even Africa. The logic of the EC (which allows for free movement across internal borders) requires uniform immigration laws, but the social pressures on governments are mostly in favour of greater restriction. That is why Germany is attenuating its liberal policy of asylum, and why France last week announced that "zero immigration" is its goal. But too great a move towards restricted entry will create intolerable social problems in Central Europe, while too liberal a policy will provide excuses, however thin, for the kinds of attacks on foreigners that have shamed decent opinion in Germany for the past several months. The dilemma would be blunted if the economic difficulties were less obdurate: then there would be less worry about jobs within the EC, while would-be migrants might be tempted by overseas assistance to stay where they are.

Finally, there is the difficulty that even stalwart Europeans are no longer sure just what Europe is for. The narrow goal of prosperity seems no longer easily attainable, and is in any case a poor substitute for more general statements of the goals of common action. Yet the disappointments of the past several months in which Europe has singularly failed to find a basis for constructive action in what used to be Yugoslavia are merely a reminder that there will be great difficulties in deciding on a response to, say, an attenuation of the commitment of the United States to Europe during the years ahead. It seems, unsurprisingly, that there can be no European dream without a statement of its purpose.

That is why there is every reason why Europeans should balance enthusiasm for "ever closer union" with a recognition that their project will come to nothing if they attempt to force the pace unreasonably. The hope must be that the purpose of the whole European exercise can be restated in terms that will make it seem worthwhile to those whose lives will be most affected. □

Microbes ever marching

An outbreak of respiratory disease in the United States warns of the need for vigilance on microorganisms.

ON an Indian reservation in the American southwest that straddles the four corners of the states of Arizona, Colorado, New Mexico and Utah, a "mystery" illness has claimed the lives of 12 people during the past couple of weeks — and has been temporarily named URDS (for "Unknown Respiratory



Stress Syndrome"). The disease, which strikes apparently healthy people and leads to death from suffocation caused by fluid in the lungs, has now been confirmed in at least 19 cases. With the IXth international conference on AIDS just over in Berlin, it is natural that the appearance of URDS should suggest that another deadly epidemic has sprung upon us. Can that be so?

It is, of course, far too soon to know. At this stage, there is no more to go on than in the first few days after the appearance of Legionnaires' disease among war veterans gathered at a hotel in Philadelphia. (Epidemiologists, virologists, bacteriologists and Navajo Indian medicine men are said to be working around the clock at Four Corners in a search for the cause of URDS.) But the incident is a clear reminder of the warnings issued recently by a committee of the Institute of Medicine under Joshua Lederberg, former president of Rockefeller University, to the effect that human beings should not become complacent about threats from microorganisms not yet identified even by their effects.

The best guess is that URDS is caused by a variant of the rodent-borne Hantaan virus originally isolated in 1976 from the lungs of a field mouse in Korea, near the Hantaan river. Circumstantial evidence has been provided by Navajo medicine men, who have told US health workers that the Four Corners reservation has an unusually large rodent population this year because the pinion tree has borne nuts year round, apparently for only the third time this century. But Hantaan virus has been isolated from only three samples of fluids from infected patients and it would be premature to suppose that the cause of URDS has been identified.

Although hantaviruses are common in Asia, they have only recently been recognised elsewhere. But in the past few years, they have shown up in places as different as the former Yugoslavia, in the harbour of Baltimore, Maryland, and the fields of Texas in the United States. What their endemic future outside their original habitat may be is, for the time being, a matter for guesswork only. So is the route by which they have spread beyond Korea.

Only if the Four Corners outbreak is the harbinger of a more permanently established infection is it likely to seem worthwhile to investigate that question. But the general principle is that, with the passage of time and the greater movement of people and goods from one place to another, geographical barriers to the spread of infection are continually attenuating.

So what should be done? In all likelihood, the Four Corners outbreak will be brought under control. The temptation then will be to forget about the incident. But that would be shortsighted. Of necessity, it cannot be long before there is a recurrence of the same trouble somewhere else. And even if the fuss at Four Corners blows over quickly, nobody can be sure that the next outbreak will be so easily dealt with. And in all likelihood, the Four Corners outbreak will be brought under control. But this incident is a telling reminder that there are reservoirs of microorganisms to which people have not yet been exposed not merely in feral animal populations but in parts of the world that have hitherto been isolated by geography. □

Pity poor understanding

Government ministers everywhere need a crash course in understanding what science is all about.

CAMPAIGNERS for better public understanding of science (*Nature* included) often give as a reason for their sense of urgency the fact that citizens in modern democracies are increasingly required to express their views on issues with significant scientific content. How, the argument goes, can the person in the street decide whether to vote for a tax on fossil fuels if he or she does not appreciate the link between atmospheric carbon dioxide concentrations and global temperature? What is the voter to make of a proposed ban on chlorofluorocarbons, if he does not know whether they destroy stratospheric ozone or contribute to the greenhouse effect? (Regular readers will know that the answer, at least in principle, is "both".)

Real-world democrats may point out that this argument neglects the role of our elected representatives: do we not delegate to them the job of getting to grips with all the facts behind the issues, whether social, economic or scientific? Is that not what we pay them for? If so, then it is to be hoped that citizens of other countries are getting a better deal than the British, who two weeks ago heard their Secretary of State for the Environment, Mr John Gummer, mistake the ozone hole for the greenhouse effect in a political interview broadcast on the radio. In discussing a proposed tax on domestic heating fuel, he made the resounding statement, "... if we are going to do something about the ozone layer, we have to ... make the use of fuel more expensive...". Nor was this an isolated incident: in the past year, listeners to the national radio network in Britain have heard one Conservative Member of Parliament say that radio waves escape from the Earth's atmosphere through a hole in the ozone layer, and another that dogs do not have DNA. (The last comment, in a discussion of legislation to control dangerous breeds of dogs, was meant to convey the difficulty of defining a breed, as opposed to a species — but knowing what the speaker meant does not make it any less of a gaffe.)

The charitable view that these solecisms are mere slips of the tongue is, sadly, not tenable. In Gummer's interview, it is true, the previous discussion had ranged over both global warming and the ozone hole. But then why did no one else participating in the discussion — neither the interviewer, nor the two opposition politicians, who seemed all too ready to criticize the hapless minister for other errors — correct the mistake? The remark about DNA was also made in an adversarial setting, with plenty of opportunity for point-scoring if an opponent had been so minded and knowledgeable; sadly, no points were scored.

Not long ago, the US public rightly considered it scandalous that their vice-president could not spell "potato", and thought that Latin Americans speak Latin. When a scientific clanger can earn its perpetrator the same degree of opprobrium, we shall know that the 'public understanding of science' has finally been attained. □

Legislation to boost US technology takes protectionist turn in House

Washington. Legislation to improve US competitiveness by helping industry to exploit new technology is being attacked by companies it is designed to help because an amended version would exclude participation by other countries.

The National Competitiveness Bill (HR 820), which includes measures ranging from the creation of local research centres for manufacturing technology to federal grants for some types of industrial research and development, was passed last month by the US House of Representatives; a similar version (S.4) awaits approval in the Senate. But two amendments approved during debate in the House give it a decidedly protectionist flavour.

The first, offered by Representative Thomas Manton (Democrat, New York) and Cardiss Collins (Democrat, Illinois), would exclude non-US companies from participation. The change was accepted reluctantly by the bill's sponsors, led by Representative Tim Valentine (Democrat, North Carolina), as a compromise to placate Representative John Dingell (Democrat, Michigan) and to prevent the bill from being referred back to the Energy and Commerce Committee, which Dingell chairs. The second amendment, offered by Mac Collins (Republican, Georgia), would effectively exclude any US company with foreign nationals on its board.

The bill, a central component of the Clinton administration's programme to strengthen US technology, would authorize the spending of \$1.5 billion over two years on a series of local centres, research grant schemes and training initiatives to promote manufacturing technology. The National Science Foundation (NSF) and the National Institute of Standards and Technology (NIST) would administer the bulk of the programme.

But the main lobbying group for industry, the National Association of Manufacturers (NAM), has withdrawn its support for the House version. It thinks that the Manton-Collins amendment would exclude thoroughly US multinational companies such as Kodak and IBM, as well as non-US companies with strong US research operations, such as the electronics group Philips. This week in Brussels, the European Commission is considering its response to what one European diplomat in Washington described as "a clear case of a non-tariff barrier to trade".

Opponents of the Manton-Collins amendment expect to prevail in the Senate, with help from the White House. But the

outcome of a conference later this summer to reconcile differences in the two versions will depend in large part on Dingell's attitude. "We're certainly hoping to get it killed", says Bill Morin of NAM about the US-only provisions. "But it's a question of how far Dingell digs in his heels."

The Senate version is similar to the unamended House version, but also contains provisions for the development of a



Manton (at left) amends, Walker objects.

high-speed telecommunications network. The House is addressing that issue in a separate bill (see *Nature* 363, 104; 1993).

The Manton-Collins amendment requires that companies participating in schemes authorized in the bill must agree to manufacture any resulting products in the United States and to "procure parts and materials for such products from competitive US suppliers". A spokesman for Kodak says that these stipulations would exclude it and other US-owned multinational companies.

The amendment also requires the home

countries of foreign-owned companies wishing to participate to satisfy a list of criteria including the existence of a programme "equivalent to this Act" and a process for setting non-discriminatory industrial standards. The latter is seen by foreign manufacturers as a catch-all clause to exclude them.

Representative Robert Walker (Republican, Pennsylvania), the senior Republican on the Science, Space and Technology Committee, criticizes the amendment as "purely protectionist". Walker argues that the requirement for "equivalent" research support schemes in other countries is far stricter than previous laws, which require only "comparable" schemes, and could never be met. Supporters of the amendment concede that its intention is to keep out non-US companies, but say that the result will be no different than what happens abroad to US-owned companies.

The acceptance by the House of the Manton-Collins language, which is unquestionably tougher than previous provisions of this type, reflects two things. One is the long shadow cast by Dingell's committee over a bill championed by the science committee, chaired by Representative George Brown (Democrat, California). The second is the prevailing mood in the House of protectionism and even isolationism. After the amendment by Mac Collins was passed 288:127, Brown said that the current mood may foil efforts to rejoin UNESCO and could also damage other forms of international science collaboration.

Colin Macilwain

Germany wants to modify gene laws

Munich. The German government has drafted changes to its restrictive gene laws so as to speed up and simplify procedures for obtaining permission to do low-risk experiments. But the draft, which faces a long and rocky road, offers researchers no relief from the procedures required for experiments defined as having no risk (see *Nature* 359, 93; 1992).

The current laws make it extremely difficult to conduct any experiment involving recombinant DNA. The proposed changes would halve the two-month waiting period at the state level for low-risk (safety level 2) experiments, and protocols would no longer need to be approved again by the federal Commission for Biological Safety. The requirement for public hearings on produc-

tion facilities using low-risk procedures would also be removed.

But the draft law does not eliminate any of the extensive documentation required for experiments defined as having no risk. A directive from the European Communities that Germany has chosen to read literally says that even no-risk experiments must be monitored, and Germany has further complicated the process with an additional layer of bureaucracy.

The draft must first be approved by the Bundesrat (upper house), where the socialist majority is expected to weaken it. The next step, approval by the cabinet, is also subject to review by the Bundesrat. The process is expected to take at least until the end of the year.

Alison Abbott

Germany debates proper niche for blue-list institutes

Munich. Time was that the only people who knew about the German 'blue-list' institutes were those who worked in them. But since reunification their number has nearly doubled, and their place in the German research scene has become the subject of acrimonious public debate.

The blue-list institutes — the name comes from the colour of paper on which the original list was printed — were created in the

scientific force to be reckoned with.

The appearance in a recession of a fourth 'pillar of research' has precipitated rare public discontent within Germany. The establishment, including the BMFT, says that the expanded list is inappropriate and that the research groups should conform to the existing landscape. It complains that the blue-list institutes are providing a temporary shelter for displaced researchers, that their range of topics — from the natural sciences to economics and public services such as museums and libraries — is too broad to justify a society analogous to the Max Planck Society and that a disproportionate number are located in the new *Länder* (see left). Their growth is contrary to the reunification treaty, it says, which gives priority to moving research from the institutes of the former East German Academy of Sciences into the universities.

Those complaints are sour grapes, says Gerhard

Neuweiler, president of the Wissenschaftsrat, who believes that the debate is the result of budget cuts caused by the recession. He also believes that most institutes will retain their blue-list status after a five-year evaluation. "They were all judged by the Wissenschaftsrat to be doing valuable research typical of blue-list institutes", he says, "that is, basic research of supraregional interest."

In response to charges that the establishment of blue-list institutes goes against the spirit of the reunification treaty, Neuweiler says that the Wissenschaftsrat recognized the impossibility of moving research straight back into universities, which were halving their staffs and which lacked the necessary

infrastructure for research. The special WIP programme to reintegrate a thousand individuals has foundered on these same grounds (see *Nature* 362, 775; 1993).

Blue-list institutes have been criticized for their inflexibility and low turnover. Unlike Max Planck institutes, they have managed to avoid changes recommended by the Wissenschaftsrat in regular assessments because of a vigorous defence of the status quo by the *Länder* that provide half their budgets.

However, this pattern may change under a plan to expand the responsibilities of the AG-BL. It is already setting up research sections to increase interaction among institutes and would also like to see a system that guarantees a true external evaluation and that enables resources to be shifted from weaker to stronger areas.

But that would require a single pot of money, which is politically very difficult to achieve. The federal government's share comes from at least eight ministries (the research ministry, which does not accept the concept of blue-list institutes as an organization for research, is the most important), and 16 *Länder* supply the rest.

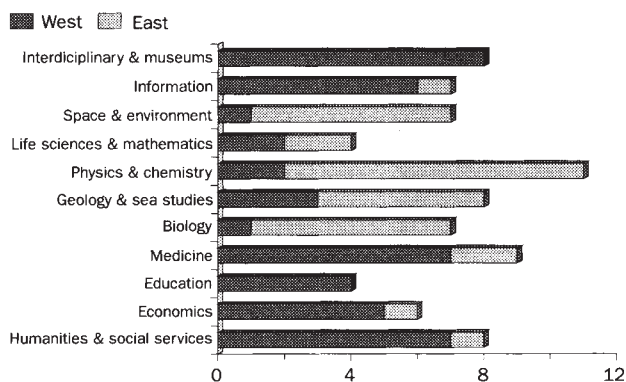
To improve efficiency within the structure, the Wissenschaftsrat is considering organizing the AG-BL along the lines of the Max Planck Society, with research sections continuously evaluating standards. But AG-BL board manager Wolfdieter Homann says that "we will be much more decentralized so that our institutes maintain their autonomy". Recommendations for the new structure will be announced next month, in time for consideration by the Bund-Land Kommission, which mediates between federal and *Länder* governments.

"In retrospect," says Neuweiler, "it is really good luck that we proposed all these institutes so early. They are the only places where science and research in the new *Länder* has been left undisturbed by the political upheavals."

Germany's major sponsor of basic research, the DFG (Deutsche Forschungsgemeinschaft), agrees that creating more blue-list institutes was the best interim solution for East German science but that the long-term goal should be university-based research.

Alison Abbott

The distribution of blue-list institutes



1970s so that the federal government could help to pay for research that was of 'supraregional importance'. The three pillars of basic research in Germany — the universities, the Max Planck Society and the national research centres — argue that the 82 institutes have distorted the way that research is carried out. Only the Fraunhofer Society, whose 48 institutes undertake applied research, has remained neutral.

In 1989 there were 48 such institutes funded equally by federal and *Länder* governments. After reunification, the German science council, the Wissenschaftsrat, evaluated all research in the former East Germany and concluded that the new *Länder* should transform three institutions into national research centres and that an additional 34 should be taken over as blue-list institutes. Following their own evaluations, the Max Planck Society established two institutes and several research groups and the Fraunhofer Society established 19 institutes and branches.

At the end of 1990, the ministry of research (BMFT) accepted the recommendations and the number of blue-list employees suddenly increased from around 5,000 to more than 9,000. In January 1991, a society (the AG-BL) was set up to look after their general interests. With a budget of DM1.16 billion (US\$700 million) — comparable to Max Planck — the institutes have become a

ESA approves new gamma-ray mission

London. The European Space Agency (ESA) has endorsed the proposal of its space science advisory committee to develop and launch a new gamma-ray observatory. Known as INTEGRAL (International Gamma-Ray Astrophysics Laboratory), the satellite will be between ten and fifty times more powerful than the two gamma-ray devices now in orbit. It was given top ranking two months ago by the advisory committee in a direct competition with three other proposed projects as the

second 'medium-sized' project in ESA's science programme (see *Nature* 363, 3; 1993).

The satellite, which is being developed jointly with the US National Aeronautics and Space Administration and the Russian Institute for Space Research, will be primarily used for observing the galactic plane and centre. It is expected to cost £400 million (US\$600 million) and is scheduled to be launched — probably using a Russian Proton launcher — in 2001.

D.D.

Internal fight threatens future of Strangeways laboratory in Britain

London. The future of an 81-year-old private British laboratory carrying out world-class research on connective tissue has been jeopardized by a disagreement between senior scientific staff and trustees over a planned

laboratory become an institute for the study of such problems.

The chairman of the trustees, Sir Barry Cross, says that the laboratory intends to ask the MRC to support a major programme on inflammation research. The council is the main financial supporter of the Strangeways laboratory which, apart from its land and buildings, lacks a significant endowment.

The other three trustees are John Bradfield, the bursar of Trinity College, Cambridge; Lord Butterfield (Peters' predecessor at the medical school); and Sir Andrew Huxley, a former president of the Royal Society. They point out that the secretary of the MRC, Dai Rees, has already told them that the laboratory must demonstrate that it can "enhance the prospect of [achieving] far-reaching conceptual advances in biomedical science" if it hopes to continue receiving substantial MRC support.

Laboratory staff dispute this conclusion, saying that it is at odds with a recent outside review that gave it alpha-plus ratings and with the MRC's renewal of five-year grants for two of its three main research groups. Others say that the staff is hobbled by the perception that connective tissue research is less fashionable than research on more visible diseases.

Senior staff who disagree with the scientific content of Lachmann's plans have been

told by the MRC that they must find other sites for their work. One group, headed by Jeremy Saklatvala, has already agreed to move to the Agricultural and Food Research Council Institute of Animal Physiology and Genetics Research in the nearby village of Babraham.

In all, more than 40 scientists and technical staff (out of a laboratory total of about 80) will have to move. The MRC says that no decision on the future of Strangeways will be made until it receives detailed research plans from Lachmann.

Strangeways staff have asked colleagues around the world to help them to block the trustees' plans. Last week, the trustees wrote to the authors of more than 70 letters that they have received, saying that they are doing their utmost to ensure a successful future for the laboratory. "Closure of the laboratory, or fundamental changes in its status, are options that we would consider only if we do not succeed in finding a way forward for it as an independent institution", the trustees explained.

The trustees and Lachmann have made it clear that they will not back away from their proposed changes in the scientific priorities of the laboratory. But many supporters of the current research programme fear that the damage has already been done.

"One of the present strengths of Strangeways is that the research group leaders have overlapping interests", says Roger Mason, professor of biochemistry at the Charing Cross and Westminster Medical School in London and chairman of the British Connective Tissue Society. "If these groups all start disappearing to other institutions, the critical mass will have been lost."

David Dickson



Strangeways laboratory in research row.

change in its scientific priorities.

The trustees of the Strangeways Research Laboratory in Cambridge, perhaps best known for its early pioneering work on tissue and organ culture, want to move into the field of inflammation research to broaden the laboratory's base and strengthen its links with the University of Cambridge. But senior scientists, many of whom are already being required to relocate their research teams as a result of their disagreement with the trustees, say that they have been excluded from discussions on the laboratory's future, and they have little confidence in what has been presented.

Opponents see the shift as a takeover by the university medical school, which is keen to strengthen its position in clinical research. The school suffers from an acute shortage of space at its Addenbrooke's Hospital site, which is five minutes from the Strangeways.

The crisis at Strangeways has been triggered by the forthcoming retirement in September of its current director, John Dingle, and the choice as his successor of Peter Lachmann, professor of immunology at the university, honorary director of the Molecular Immunopathology Laboratory of the Medical Research Council (MRC) and president of the Royal College of Pathologists. The trustees have endorsed Lachmann's proposal to concentrate on the mechanisms of inflammation and one of the laboratory's governors, Keith Peters, regius professor of physic (medicine) at the medical school, says that he would eventually like to see the

Swiss loosen restrictions on foreign workers

Basel. Switzerland is easing its historic restrictions on foreign workers, making it easier for researchers to get a Swiss work permit and to leave the country temporarily. One measure of success will be the extent to which research institutes will be able to recruit technical staff from abroad.

The changes loosen decades of restrictive policies on foreign industrial workers, under which each of the country's 26 cantons are allotted a certain number of permits and employers have to prove — through advertisements — that no Swiss national is qualified for the job. But facing a deepening recession, Switzerland decided that, according to a government spokesman, "restriction in the exchange of people might hamper recovery". Last autumn Swiss voters passed a referendum delaying membership in the European Economic Area (EEA) (see *Nature* 363, 8; 1993), and politicians have reacted by trying to show that Switzerland is "by no means an isolated island in Europe".

The new law, which went into effect on

1 May, does not increase the quota of foreigners, but it eliminates the requirement to advertise vacancies locally. It also allows foreigners to leave the country to take courses without having to reapply for a work permit.

The changes are welcomed by the chemical and pharmaceutical industries, which make up 21 per cent of the Swiss economy. "The new regulations make recruitment of personnel much easier", says Heiko Jürgensen of Ciba-Geigy.

Many of the 10,000 foreign scientists working in Switzerland are employed by Basel's 'big three' pharmaceutical companies — Ciba-Geigy, Sandoz and Hoffmann-La Roche — and commute to work from France and Germany. Basel and other cantons close to the borders have special quotas that allow for additional *Grenzgänger* (border crossers).

Although these restrictions have never applied to academics, laws such as a ban on buying property make life uncomfortable for foreigners. And officials are always looking for violators.

Oliver Klaffke

Biology in China suffers from government neglect

Beijing. The life sciences in China are being left out in the cold, say a group of prominent researchers who have asked the government to create a national biomedical research fund. As evidence, the scientists point to a growing gap between China and the rest of the world in such fields as molecular biology, genetics and immunology as well as a recent decision by the State Education Commission to eliminate biology from the national college entrance examinations.

"The progress and development of biomedical science has direct implications for controlling population growth, improving living standards, and mitigating the effects of ageing and disease", according to a recent letter to the State Science and Technology Commission and to the Ministry of Public Health signed by 10 well-known biomedical scientists. "All developed countries have placed biomedical sciences in strategic po-

sitions and have set up a special fund to provide long-term consistent support."

The scientists, including Wu Jie-Ping, a member of the Chinese Academy of Sciences, lament the size of the Chinese biomedical community and its output relative to other countries, including those in the developing world. A recent analysis by the Institute for Scientific Information in Philadelphia, for example, found that the relative impact of Chinese papers in molecular biology, genetics, cell biology and immunology has fallen substantially in the past five years, whereas research in metallurgy, computer sciences, nuclear engineering and instrumentation has become more visible.

The government's attitude is reflected in its decision last year to exclude biology from the university entrance examinations. The proposal was part of a reform plan submitted by the Beijing municipal educa-

tion authorities to the state commission, which quickly accepted it. The changes, which take effect this year, have been adopted by 10 other provinces.

There are two versions of the current tests, one for science majors and one for social science students. While everyone must be tested on Chinese language, mathematics, politics and English, the science examination also includes physics, chemistry and biology and the social sciences covers history and geography.

The official reason for the change, which also drops geography from the list of subjects to be covered, is the desire to reduce the work-load on high school students, who compete intensely for the chance to be among the 3 per cent of the population admitted to institutions of higher education. But biologists and geographers believe that the strain on students should be relieved by improved teaching methods and better textbooks, not by narrowing the scope of their education. They also fear that high schools will start to drop the two subjects from their curricula if students do not need them for university admission.

Biologists are also worried that introductory courses will have to be simplified if students have not taken the subject beforehand and that, as a result, students will learn less biology during their undergraduate years. As an alternative, the researchers have suggested combining physics, chemistry and biology into a single test. **You Qin Li**

Canadian drug companies fund universities

Quebec. Canada's leading pharmaceutical companies have increased their support for a government-run partnership as part of a promise to spend C\$200 million (US\$160 million) over the next five years on university-based research. The money will be distributed by the Medical Research Council (MRC), the chief federal source of biomedical research funding.

The action by the Pharmaceutical Manufacturers Association of Canada (PMAC) comes at a time when the Canadian government has told researchers to look to the private sector or to their provincial governments for additional funding. It is consistent with the MRC's new strategic plan, which places great emphasis on such partnerships (see *Nature* 361, 102; 1993). And it follows on federal legislation enacted last year that extends drug patent protection in return for greater research spending by industry (see *Nature* 355, 666; 1992).

The partnership enlarges a programme begun in 1986 in which industry provides two dollars for each dollar spent by the MRC. The grants are for industry-related research at universities as well as for fellowships and visiting chairs to attract experienced industrial scientists to campus. The government intends to increase its annual contribution to the programme from C\$7 million to C\$10 million, thereby raising industry's share to C\$20 million a year.

The money will also set up a two-way, national electronic database to provide industry with greater access to those in the MRC's peer-review system and to other academic scientists. The current systems are

university-based, limited in scope and not accessible by industry. "It's a little ponderous and it takes a long time to work through", says Gordon Postlewaite of the PMAC. "The whole programme is being restructured by MRC to be much more responsive."

Using the new system, a company will be able to identify experts in a particular field; in return, universities can increase their contacts with industry. The database provides industry with what Postlewaite calls "the Good Housekeeping stamp of approval — the high quality of MRC peer review — and a very efficient turnaround time". The partnership will pay start-up costs for the database, which will later be financed with user fees.

The partnership gives industry a chance to train researchers in specialized niches with potential labour shortages, while providing university researchers with additional funding. Intellectual property rights to inventions arising from such joint activities will be settled through negotiations.

Although Postlewaite admits that some researchers may be concerned about accepting money from industry, he says that there is no attempt to hide the funding source and that, in any event, such reservations are no longer valid. "A few people in basic research still have this expectation that industry somehow owes them money to spend as they see fit, without any accountability or relevance to the marketplace's need for a return on investment", he says. "But with money as scarce as it is, I think that most universities have decided to face reality."

David Spurgeon

Storm damages largest US station in Antarctica

Washington. A two-day storm last week with sustained winds of more than 100 knots damaged several buildings at McMurdo Station, the hub for US activities in Antarctica.

The punishing storm ripped off the roof of a building containing high-frequency radio transmission equipment and embedded a sheet of plywood from a storage shed into the wall of another building several hundreds yards away. No serious injuries were reported among the 236 people wintering over at the base, which has a summer population of about 1,200.

Officials trying to assess the damage are hindered by snow drifts as high as 12 feet produced by snow that blew horizontally for 48 hours. David Bresnahan, station manager, says that the repairs will take several months and that construction crews will have to be diverted from other scheduled activities, including the completion of the new science building. The loss of the high-frequency radio equipment, he adds, is likely to hinder communications next season with scientists at field stations in the interior of the continent. **J.M.**

A hopeful Chile looks for ways to channel its advantages into science

Santiago. Chile dreams of becoming a world-class power in science. The country has many advantages, among them a vibrant economy and extraordinary international facilities in astronomy. It also has a highly educated population and strong ties with the global scientific community.

recall the conflicts this situation creates. Warnings from their US or European mentors that good work would be nearly impossible in Chile are usually accurate, but they also sense an obligation to their country to contribute what skills they possess.

There is no history of postdoctoral training in Chile; next year, the government will award its first postdoctoral scholarships to 20 locally trained scientists as a way of enlarging existing research

not to give tacit approval to the coup that deposed the popularly elected leftist government of Salvador Allende.

But the government in recent months has taken steps to show that it believes science and technology are vital to the country's future. Scientists, optimistic that some of their wishes may in part be fulfilled, have responded by asking for even greater support.

The government's most visible action — and possibly the most controversial — was its decision last autumn to buy a 5 per cent share of a \$176-million project to build two 8-metre international telescopes, one in Hawaii and the second in northern Chile. The United States, Britain and Canada have pledged the bulk of the money for the project, called Gemini. Brazil and Argentina have recently agreed to contribute the remaining 5 per cent.

There is a consensus that Chilean science should receive a tangible benefit for providing foreign research organizations with the land and logistical support to build some of the world's most powerful astronomical observatories. But there is disagreement on what that should be (see page 488).

The tiny community of Chilean astronomers — only about two dozen scientists require time on modern telescopes — see their government's commitment to Gemini as the first step towards providing the infrastructure needed to build up a world-class astronomy programme. But others hold that the US\$9 million is a disproportionate investment by a government that spends only \$20 million a year on basic research; they would prefer the money to have been spent on attracting more students into science or on stimulating investment in research by the private sector.

In another promising development, the government has recently awarded the first series of grants from a four-year, US\$65-million initiative to put academic scientists to work on projects with a tangible economic return. The money comes from a programme operated in parallel with one of similar size that funds basic research.

The new fund for scientific and technological development, known as Fondef (see page 486), was created to comply with the terms of a US\$100 million loan from the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CHILE'S BOUNTY: Aerial views of (top) Cerro Tololo Inter-American Observatory, (right) the European Southern Observatory and (below) a computerized image of the Gemini telescope at Cerro Pachon.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

But it is not easy to be a scientist in Chile. Salaries for university professors are inadequate. There is a dearth of opportunities for those trained in research. So those already trained are held back, and students are discouraged from entering the profession. The relatively small size of the community — Chile has an estimated 2,000 active researchers in a population of 15 million — means that there are not enough people to train the next generation of scientists. Chile awards only about 20 PhDs in science each year; most of those seeking professional training go abroad.

And they tend to stay there. Many of the best research graduates remain in their new homes because opportunities in science are better there. Prominent Chilean scientists

groups and increasing the system's overall capacity. Prospects for new scientists in industry are no better. Chilean industries rarely hire PhDs, and the idea of working for industry is still frowned upon in academic circles. The applied research taking place in Chile is tied closely to the abundant natural resources that underpin the national economy. There are 16 state-supported research institutes covering such areas as mining, forestry, fisheries and agriculture, but they work mostly on contract for individual companies.

To be sure, the military dictatorship of General Augusto Pinochet, from 1973 to 1990, dealt science a crippling blow. The size of faculties at some universities shrank by half, and money for equipment, journals, travel and other ways of keeping up with international science all but disappeared during the first years of the dictatorship. A further obstacle was the ostracization by the rest of the world, anxious

This supplement was written by Jeffrey Mervis. *Nature* would like to thank all those who helped with arrangements, in particular the Chilean Academy of Sciences for its cooperation and hospitality.

World Bank. Universities and non-profit organizations compete for grants averaging more than \$US1 million over three years, extremely generous by Chilean standards. Some 45 awards were made in the first round from a pool of 200 applicants.

Teams of university researchers will work with companies on a practical problem, and the government expects them to succeed. Applicants were required to promise that their projects will generate a return of at least 10 per cent on the government's investment; in the next round, applicants

must find industrial sponsors willing to contribute at least 15 per cent of what the government is spending.

How much science can Chile perform? There is hope that Chile's strong economy will lead to greater government spending on university-based research as well as to greater opportunities in the private sector. And improved regional cooperation would help it to overcome the limits to growth imposed by a small domestic market. In the meantime, Chilean scientists are striving for excellence in whatever they do. □

Spending increases do not match the need for jobs and better pay

Santiago. Government support for science in Chile has increased substantially in the past decade, but not fast enough to satisfy Enrique D'Etigny, president of the Comisión

been able to place them in industry", he says. As a result, he says, "students do not see science as an attractive future."

Fondef is intended to change that perception by showing that university scientists can contribute to economic development. But D'Etigny recognizes that the professors must also change.

Cell biologist Bernardo Gonzalez, an assistant professor at Católica University with a Fondef grant for a bio-bleaching project involving the pulp paper industry, knows what the pressures are. The grant, US\$1.5 million over three years, is one of the biggest in the biology department, but he says that several senior faculty members opposed his application because it would

bring the department into direct contact with industry.

Gonzalez, who is 33, wants to show that there is more to biochemistry than teaching

and publishing. He says that he enjoys his collaboration with Chile's second largest pulp company, Celulose Arauco, and that he will keep his ties to the "outside world", including a possible stint abroad with a company, even if it makes it harder for him to obtain tenure.

Industrial support also provides funds for undergraduates, which the typical government grants do not. Gonzalez hopes that this may offset the feeling among good high school students that they cannot make a living as scientists, and who go into medicine or law instead.

Part of the crisis within Chilean universities stems from a crackdown early in the 17-year reign of General Augusto Pinochet, which purged departments, reallocated resources and decentralized the university system to weaken the political clout of the University of Chile. (The number of universities rose from eight to 64 in a dozen years, including 22 that had been branch campuses of the University of Chile.) The ending of free tuition was accompanied by a sharp decline during the 1980s in government support for higher education, which is now less than half the total university budget.

The approach to science by the three-year-old Aylwin government, whose successor will be chosen in December, mirrors that towards the booming Chilean economy as a whole: make use of the country's 'comparative advantages'. But scientists emphasize that the government, while supporting aquaculture, mining, agriculture, forestry and the like, must not forget that progress in such areas rests upon a healthy research base.

D'Etigny says that his goal is to double, to 1 per cent, the share of Chile's gross domestic product that the government spends on research and development (see below) and also to double the output of science students. Following the lead of the Andes Foundation (see page 487), Conicyt has created a programme of doctoral and postdoctoral fellowships offered only at



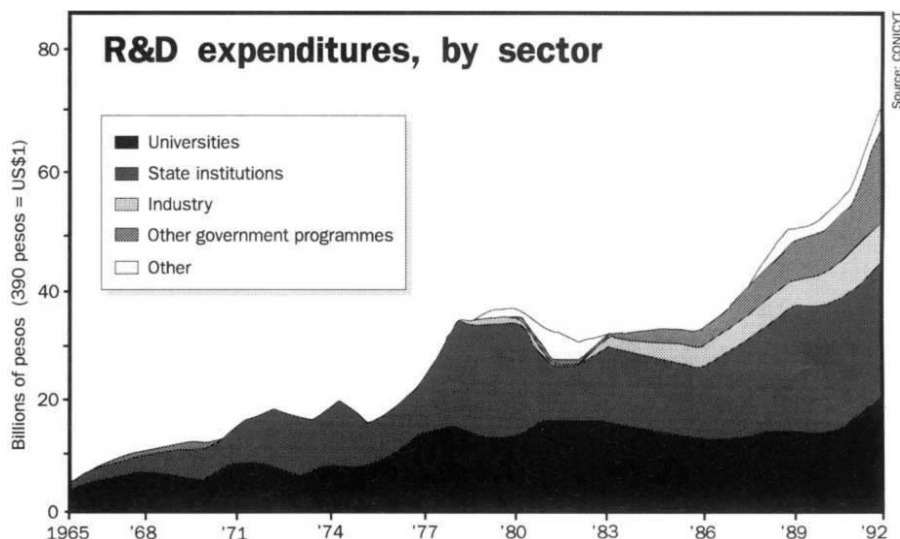
Jeffrey Mervis

Enrique D'Etigny is Chile's *de facto* science minister.

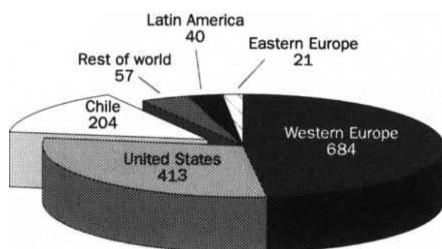
Nacional de Investigación Científica y Tecnológica (Conicyt). As head of the equivalent of the US National Science Foundation as well as science adviser to Chile's president, Patricio Aylwin, D'Etigny says that he would like the research system to expand until it is able to attract the most talented undergraduates and to support those who have been trained for scientific careers.

Conicyt, created in 1967, operates a peer-reviewed grants program for university researchers called Fondecyt (National Fund for Scientific and Technological Development) as well as a newer programme, called Fondef, that provides grants to academic research offering a commercial return. Its overall budget of US\$40 million may seem small but there are only about 2,000 active Chilean scientists with PhDs, only 15 per cent of whom were educated at home (see table next page).

D'Etigny says that basic science could grow "probably by 20 per cent" to meet existing demand, but that further expansion would require a broader base. "The salaries of young scientists are low, and we have not



Where Chilean scientists get their PhDs*



*Fondecyt Investigators, 1989–1991
Source: CONICYT

departments whose programmes have passed a national evaluation.

But he is frustrated by his inability to influence the way in which universities spend the lump sums they receive from the education ministry. "Some spend a lot of it on research, and others allocate nothing", he says. A new Conicyt programme that provides researchers with a 15 per cent supplement to their salaries helps to ease the pain, but it is not enough to make a real difference.

D'Etiigny would also like to improve communication between the government and university researchers, including greater input from outside advisory panels. "We don't have a good way to develop new areas of science", he says. "That's something that needs to be worked on." □

graduate scholarships in science should not be squandered on substandard programmes. So the foundation sponsored a comprehensive review by British scientists of Chilean science departments: the scholarships awarded were then reserved for students at departments that passed muster. Three years ago, the foundation repeated the exercise with a panel of US scientists.

Another perennial problem for Chilean scientists is the low pay of university professors. Recognizing that talented students were being attracted into the more lucrative fields of medicine, business and the law, the foundation set the annual stipend for its scholarships at a level equivalent to what some university lecturers were earning. Saavedra has little sympathy for his colleagues who complain, telling them "the problem is not that we are paying too much, but that our universities are paying too little".

Although the foundation could afford only eight four-year scholarships a year, the awards proved so popular that two years ago the national government began a larger programme of its own. Since then the foundation has begun a more selective programme — one scholarship each in the fields of biology, chemistry, physics and mathematics — with an even higher stipend, again following Saavedra's preference for quality over quantity. And in 1990 it began a programme of postdoctoral awards for students to complete their education overseas, usually in Europe or the United States.

The foundation also makes small awards to individual researchers for work relevant to industry or to groups needing special equipment, and it has supported a growing academic computer network linking Chile with the rest of the world. It hopes to increase opportunities for young scientists by connecting new and emerging universities to researchers at established universities. And last year it began a joint research programme with its sister foundations by awarding 20 grants for projects involving scientists in Argentina and Brazil.

The foundation's most recent, and in many ways most ambitious, project is a US\$9 million technical high school in Concepcion, about 500 miles south of Santiago, to encourage students with a more practical bent. Modelled on schools in Israel sponsored by the Organization for Rehabilitation and Training, the school will offer disadvantaged teenagers a four-year technical curriculum, free of charge, as well as retraining for older workers. And the foundation hopes to join with industry to upgrade the curricula of 30 other technical schools around the country.

Although the foundation is gradually moving towards greater support for social and cultural projects, Saavedra does not expect its role in fostering science to change. "We're small", he says, "but we want to have an impact beyond our means. And the best way to do that is to insist on quality in whatever we do." □

Unique fund targets quality

Santiago. Money from a foundation created by a Latin American mining magnate has made possible a series of innovative programmes to bolster science and technology in Chile.

In a country without a history of private philanthropy, the Andes Foundation has been one of the most positive developments in Chilean science of the past decade. Its accomplishments range from the creation of a programme of graduate fellowships that has since been copied by the national government to the opening next summer of a technical high school that will give disadvantaged students a chance to receive first-

social welfare — a mandate that allows for considerable flexibility. The fact that the Andes Foundation devotes more than 60 per cent of its budget to education, most of it targeted at basic academic research, reflects the influence of Igor Saavedra, a theoretical physicist at the University of Chile and a member of the foundation's six-member board of directors.

"They came to me because they wanted someone from the university", Saavedra says about a steering committee formed by Stan Davidovich, a prosperous engineer and a director of the company, and Sergio Marksmann, a successful entrepreneur. "They

"Chile needs an intellectual élite to survive and flourish. I hope the foundation can play a role in its creation."

Igor Saavedra
Andes Foundation trustee



Jeffrey Mervis

class training in electronics, computers and other high-technology fields.

The Andes Foundation is one of four funds created in 1986 from an endowment of US\$250 million left by the dissolution of a consortium of mining companies headed by Luis Hochschild. On Hochschild's death, company officials sold the business, which operated throughout Latin America, placing the assets in a Lichtenstein foundation. The income from the endowment is divided among three other foundations serving Argentina, Brazil and Chile; each receives about US\$5 million a year.

Each foundation is required to spend its money improving educational, cultural and

hadn't decided how to spend the money, and they were willing to listen to my suggestions." Over the course of several meetings, Saavedra persuaded the group to set high standards and maintain them at all costs.

The results have been programmes devoted largely to attracting talent into science and expanding opportunities for those already in the field. "Chile needs an intellectual élite to survive and flourish", says Saavedra. "I hope the foundation can play a role in its creation."

Saavedra went to Britain for his graduate training and so knew how few were opportunities at home. But Saavedra has nevertheless insisted that the creation of the first-ever

Astronomers hunger for more despite richness around them

Santiago. The most powerful telescopes in the Southern Hemisphere peer out at dry, cloudless skies over the mountains of northern Chile. But when the astronomers who use them pack up and head home, only a tiny handful remains in Chile. Most board aircraft for Europe, the United States and Asia, like modern-day pirates returning with their treasure of scientific observations.

With only a few dozen active astronomers, Chile cannot be a leader in the field. But it is no longer content to be merely the home of world-class astronomy — in particular, the European Southern Observatory (ESO) at La Silla, the Cerro Tololo Inter-American Observatory (CTIO), operated by the US-based Association of Universities for Research in Astronomy, and the Carnegie Institution's Las Campanas observatory. It wants a larger share — more observing time, more cooperative programmes and more public benefits as payment for providing the sites (see *Nature* 363, 384; 1993).

Two things must happen first: the world must acknowledge Chile's contribution and Chile must put more of its own resources into astronomy to take advantage of those opportunities. Alone, neither is sufficient.

"If the observatories were elsewhere — in a European country, for example — the problem wouldn't exist", says Bob Williams, who is leaving Chile in August

after eight years as CTIO director to become head of the Space Telescope Science Institute in Baltimore, Maryland. "It would simply be accepted that the host country deserves a share of viewing time. But if this were a mining enterprise rather than an observatory, the government would recognize its value and invest accordingly."

CTIO built into its original agreement for the 4-m telescope 10 per cent of the viewing time — 18 days in each six-month schedule — for astronomers at the University of Chile, regardless of the quality of their proposals. Williams admits that a few of the successful Chilean proposals are weaker than some rejected proposals from elsewhere, but he says that happens "maybe four nights out of 180, and only on the large telescope, and that the overall impact is minuscule." And he adds, "funding that work really benefits science in Chile".

By contrast, ESO allocates time strictly on the basis of quality (although its decisions yield a distribution that closely matches the financial contribution of each member country), and the leading Chilean astronomers are funded sporadically, if at all. That has led to tense relations between Chile and ESO, and is one reason why the government is seeking a guaranteed minimum amount of time on the Very Large Telescope being built on a second site, some 300 km north of La Silla.

ESO believes that Chile's request for a 10 per cent share is unreasonable and that it should instead strengthen its university programmes so that its scientists can compete internationally. "Holland has a community five times larger and it gets only about 7 per cent of the time", says Jorge Melnick, a Chilean astronomer trained in the United States who is director of the La Silla observatory. "And what are Chilenos going to do with that time, anyway? What Chile needs is to invest in people — find good students and offer them jobs with decent pay after they are trained."

Chilean astronomers admit to falling well short of that goal. "We have only a master's programme", says Maria-Tereza Ruiz of the University of Chile, "and we don't want a PhD programme yet because we don't think we have enough people here to train world-class astronomers." But she and her colleagues believe that more observing time is essential to that goal.

"Astrophysics is one of the few sciences

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Jorge Melnick at ESO's guesthouse in Santiago.

in Chile that has the potential to be at the frontiers of knowledge", she says. "But we wouldn't be here if they [CTIO] weren't here and allowing us to have 10 per cent. At ESO, they don't care about training the next generation of Chilean astronomers."

Melnick disagrees, pointing to donations of computer equipment, journal subscriptions, vehicles and other material to several Chilean universities. He says that ESO has invited Chile to become a member, thereby giving Chilean technicians a status equal to Europeans (who are paid much more as an inducement to live in Chile) and Chile a greater voice in organizational matters, but that "all they care about is telescope time". He says that ESO "will be happy to help" once Chile "decides how much it wants to commit to astronomy".

Gaspar Galaz, a graduate student in astronomy at the University of Chile, is also eager for an answer from the government. He is living proof of the potential impact of the international observatories on Chilean science, as well as the tenuous position that astronomy holds in any hierarchy of possible careers for talented students.

"I remember visiting Cerro Tololo as a teenager, with my family, and thinking, 'Gee, I'd like to do that someday.' If I had grown up in Venezuela or Paraguay, I'd probably be studying something else." □

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

U. of Chile astronomers José Maza and Maria-Tereza Ruiz analyse red shifts in quasars.

Biotechnology looks for niche that will profit local science

Santiago. Chilean-born Pablo Valenzuela had learned the risks of starting a biotechnology company as a co-founder in 1980 of the US-based Chiron. But it was still a big step when in 1986 he and two other Chilean scientists from Catolica University decided to start the first one in their own country.

Seven years later, Bios Chile is alive and well. In January, the company and its 28 employees moved into larger offices to accommodate its steady growth. Arturo Yudalevitch, another co-founder, has greatly reduced his university links to spend more time managing its research programme.

Having proved that biotechnology can take root in Chile, the founders of Bios Chile have set their sights on a more ambitious goal: to create a successful Latin American pharmaceutical company with strong links to university-based research that can be a model for biotechnology and other high-technology sectors of the economy. Their reward, apart from money, would be a self-supporting research base and greater opportunities for scientists.

"The big problem with Chile is that we produce scientists and they all end up elsewhere", says Valenzuela, senior vice president for research and development at Chiron, in Emeryville, California. Valenzuela returns to Chile ten times a year for several days at a time. He says that "there is no industrial biology to speak of, so there's no place for new PhDs to get a job. One of things we hope to do with Bios Chile is to show that a small company can offer an alternative to the university."

Bios Chile is following a conservative business plan that reflects the risks of charting a new course without government support. It sells custom-made monoclonal antibodies to scientists in the United States, Europe and Japan, and last year it acquired Prater Laboratories, a Chilean company with 300 employees and annual sales of \$7 million from generic drugs and cosmetics. At the end of 1991, Chiron bought 19 per cent of the company's stock and a subsidiary called Austral Biologicals was formed in San Ramon, California, to sell a range of specialized biological products made by Bios Chile, Chiron and others. The company also has a joint venture with a Belgian company, Eruogentec, to develop a full line of products for the salmon farming industry. (Chile is now the world's second largest exporter of salmon, behind Norway, after recently surpassing Scotland.) Its scientists are working on a vaccine against hepatitis B and on various diagnostic products for

pregnancy and blood testing.

The company's founders do not expect Bios Chile to compete with global drug companies and their vast resources. But although they are not looking for a handout,

monoclonal antibodies and some other niches, we can grow and prosper."

Although most of Yudalevitch's academic colleagues still look down their noses at collaborating with industry, he says that a handful of graduate students are working on various company projects. There are only half a dozen biology departments in Chile of sufficient size and quality to make such collaborations possible.

Yudalevitch never had that opportunity. As an undergraduate during the 1960s, he had to travel abroad, in his case to Britain, for his graduate training. Despite many job offers from US institutions, he accepted a position at Catolica University in 1970. Within a few years, however, working conditions under the military dictatorship became so bad that he accepted a fellowship at Albert Einstein Medical College in New York City, where he worked under Jerald Hurwitz and next door to Valenzuela.

In 1975, with his leave about to expire, Yudalevitch had to decide whether to remain in New York or return to Chile. Although he recalls that Hurwitz told him "you're wasting your time [trying to do meaningful science] in Chile", Yudalevitch chose to return to Catolica University, where, 18 years later, he is winding down his research to devote himself to Bios Chile. "You can do good work in Chile", he says, "but it's not easy."

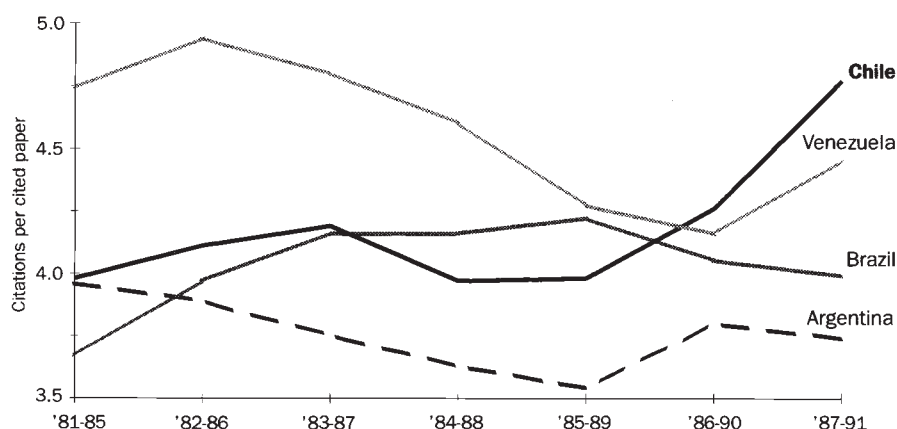
Although success is uncertain, Valenzuela is already planning for the time when Bios Chile has ten times its current number of employees and occupies an important place in the global biotechnology industry. "I'm not planning to stay in the United States forever", he says. "When I left [Chile], it was because of the military government, and I've always intended to go back. But the company needs to grow a bit before there's room." □

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Yudalevitch outside Bios Chile's new headquarters.

they do not want to be completely on their own. "We need an umbilical cord with the First World", says Valenzuela, "because Chile is isolated. But in this business, unless the government or a big foundation is behind you, your primary aim must be survival. We think that by focusing on

Chilean biology rises to the top



Source: Institute for Scientific Information

Social status and sex

SIR — R. A. Beck's letter¹ about artists' offspring provides interesting information, but fails to supply an appropriate theoretical framework against which his empirical findings could be tested. Beck's findings would have been much more instructive if he had explained why artists listed in the *Who's Who in Art* and *The Annual Obituary* might have more sons than daughters.

The appropriate framework would be the premises and the prediction of the Trivers-Willard hypothesis²:

(1) As there is such a large difference in investment necessary for a live birth between males and females, variance in reproduction should be higher in males: more males should have more children than the median number and more males should remain childless than females.

(2) To the degree that parents' access to material resources positively affects the survival chances of offspring, parents' status should positively affect their reproductive success in relation to the population average. Then, because of males' greater variance in reproductive success, high status males should be reproductively more successful than high status females, while low status males should be reproductively less successful than low status females.

(3) The prediction is that under these conditions evolution should favour a bias in the sex ratio of offspring towards the male sex in high-status individuals, and a bias towards the female sex in low status individuals.

Whereas premises and prediction have been repeatedly verified for non-human populations, evidence for human populations is rare, not least because in census data children are usually documented only with their mothers.

Here are some other examples of male-biased sex ratio among élites. In a dataset of 1,014 male members of the US élite, born between 1860 and 1939, drawn at random from *American Who's Who*, of those who gave complete information on their children's sex, there were 1,180 sons and 1,064 daughters, a sex ratio of 1.109, still differing from the US average of 1.06 for this period ($P < 0.005$). In another dataset of 1,757 male members of the German élite, born between 1830 and 1939, drawn at random from the German *Who's Who*, of those who gave complete information on their children's sex, there were 1,473 sons and 1,294 daughters, a sex ratio of 1.138, which differs from the mean sex ratio of all live births in Germany 1872–1990 of 1.0512 ($P < 0.001$). (A frequent objection against sex ratio observations in élite samples refers to an eventual underreporting of daughters. Here, from

the same sources, a dataset of 1,628 female members of the German élite, born between 1860 and 1945, was collected. The proportion of people who had not given complete information on their children's sex, was the same (about 25 per cent), as also the sex ratio of offspring among these élite females was the same (1.146) as among male élite members — the latter in conformity with the theory. It is therefore unlikely that there was a greater tendency to conceal a child's sex, if it was female.

Now, the Trivers-Willard hypothesis, strictly speaking, refers to whole population only insofar as they are single unsegmented marriage markets. Individuals who are considered of high status in the society as a whole may actually be low status on their proper marriage markets. The theory would then predict an unbiased or a female-biased sex ratio; this is what Boone³ found for the untitled gentry in the late mediaeval/early modern Portuguese genealogies.

From Great Britain there are two interesting observations with respect to this. First, fecundity data were obtained from biographies of 1,179 eminent British industrialists⁴ born between 1789 and 1925, of whom only very few had a college education and almost none went to leading universities. These men and their families certainly did not form a closed marriage market, but belonged to a much larger middle-class marriage market, in which they may have been attractive competitors. The sex ratio of the children (1,789 sons and 1,522 daughters) is 1.1754, well above the British average of 1.06 ($P < 0.001$).

Second, fecundity data were collected for 325 male members (60 per cent with a hereditary title) of the British élite, born between 1860 and 1939, drawn at random from the British *Who's Who*, whose father was mentioned in the same or earlier volumes of this source. For such men, admission to *Who's Who* was easier than for the rest; they may well be or have been less energetic and less forceful personalities than élite members without a prominent father. Of those who gave complete information on their children's sex, there were 241 sons and 249 daughters, a sex ratio of 0.968, well below the average ($P < 0.05$).

Ulrich Mueller

ZUMA, Postfach 122155,
D-6800 Mannheim 1, Germany

1. Beck, R. A. *Nature* **356**, 189 (1992).

2. Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).

3. Boone, J. L. *Am. Anthropologist* **88**, 859–878 (1986).

4. Jeremy, D. J. (1984–86) (ed.) *Dictionary of Business Biography. A Biographical Dictionary of Business Leaders Active in Britain 1860–1980*. Vols 1–5 (Butterworths, London, 1984–86).

Molecular deconstructivism

SIR — Your readers should be aware of a cruel joke at the expense of the vocal chords of molecular biologists and the integrity of the English language that is currently being perpetrated by two otherwise respectable scientists¹. In a paper in the official organ of the US National Academy of Sciences, Jürgen Brosius and Stephen Jay Gould propose to replace a large number of perfectly acceptable terms in molecular biology with a new “nomenclature” that will “keep pace with the levels and depth of analyzing and understanding genomic structure, function and evolution”.

The basic element in their new dictionary is “nuon” (not to be confused with muon), which is supposed to denote “any stretch of nucleic acid sequence that may be identifiable by any criterion”. (Note that this term does not cover invisible, nonexistent or imaginary DNA and RNA sequences.) In a clear fit of physics envy, Brosius and Gould coin such linguistic novelties as “naptonuons”, “xaptonuons”, “potonuons”, “aptonuons” and “retropotonuons”. Fortunately, for some unclear reasons, these authors decided not to rename all existing genetic elements, so that we were spared such words as “promonuon” (for *promoter nuon*), *fipriflanuon* (for *five prime flanking nuon*), or *clatwinternuons* (for *class two intervening nuons*). However, the paper does contain several nuonless novelties, such as *potomass* (for the obviously unsatisfactory genome) and *xaptoproteins* (for multifunctional proteins). One should also add that one of these authors is the co-inventor of such useful gems as “aptation”, “exaptation” and “nonaptation”², to which Brosius and Gould now add “potaptation”. (I knew all along that something was missing.)

The story gets even more complicated since a potonuon can turn in time into either a naptonuon or a xaptonuon, and we therefore have “naptonuous potonuons” and “xaptonuous potonuons”. I can just imagine myself lecturing to my students on the *coxII* genes of legumes in this Newspeak: “The pea potonuon, although clearly poto-exaptive, ended up as a naptonuon because it could not be coopted into xaptonuonity.” In the immortal words of Queen Victoria, “We are not amused.”

Dan Graur

Department of Zoology,
Tel Aviv University,
Ramat Aviv 69978, Israel

1. Brosius, J. & Gould, S. J. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10706–10710 (1992).

2. Gould, S. J. & Vrba, E. *Paleobiology* **8**, 4–15 (1982).

Contaminated blood case nears end

The trial in Paris of four scientists for negligence in the protection of haemophiliacs from infection with HIV in the period 1983–85 will cast a long shadow over the links between science and the public service everywhere.

Paris. After a month of hearings, the re-run of the great contaminated-blood trial is nearing its end, perhaps today (17 June). The lawyers hope, perhaps unreasonably, that by the end of the first week in July, there will be a second verdict on the guilt or innocence of the four scientists accused of having allowed blood products contaminated with HIV onto the French market during the first nine months of 1985. By then (but even earlier), techniques were available for preventing that state of affairs. In France, the most conspicuous victims were those then newly diagnosed with haemophilia, whose relatives have filled most of the public benches of No.13 Court in the Palais de Justice for the past month.

It is a curious setting for such a drama. No.13 Court might be a slightly shabby school-hall were it not for the high bench at one end behind which the three appeal judges sit. (There is also a reserve judge, in case another should become sick, and two prosecutors, who appear to say nothing.) The four defendants sit to the Court's left hand in a line perpendicular to the judges' bench, and behind their lawyers. Although the legal people are distinguishable by their mostly shabby gowns (no wigs), others take off their jackets when the temperature rises.

The chief defendant, Dr Michel Garretta, who became director of the National Centre for Blood Transfusion (CNTS) in October 1984, is slightly built and is made to seem even slighter by the gendarmes who sit on either side of him, their pill-box hats giving them an extra 10 cm or so; they are there because Garretta travels each day from the prison at which he began serving his four-year sentence last October. He, no doubt, will be sorry when the trial ends.

But there is very little communication between him and his co-defendants. No longer are they close colleagues. Indeed, part of the reason why they are now in the dock is that they seem never to have been close. Last Wednesday, Garretta sat with hands folded listening attentively to a remarkable histrionic performance by Maître Olivier Schnerb, the lawyer representing Dr Jean-Pierre Allain, previously director of research at CNTS, but now a professor in the department of haematology at the University of Cambridge. Allain was sentenced to four years' imprisonment, two of them suspended, last October; he appealed within a day.

The circumstances of this second trial are curious in themselves. After two appeals, the court elected to hear the whole case again. The other defendants are Professor Jacques Roux, in 1985 director-general of health in the National Ministry of Health (given a four-year suspended sentence last year) and Dr Robert Netta, director-general of the National Laboratory of Health (who was acquitted). In principle, the retrial could bring more severe sentences for some of the defendants. In practice, the outcome is likely to be leniency.

That the affair of the contaminated blood is a scandal cannot be denied, and is not. So much has been crystal clear since the publication in September 1991 of a report by Michel Lucas, the inspector-general of social affairs, whose report amply documented the reasons why, between the spring of 1984 (when HIV was well characterized serologically) and 1 October 1985, the steps that could have been taken to protect patients with haemophilia from blood products derived from contaminated blood were not taken. The CNTS, given the status of a private foundation a decade earlier but still controlled by the government, and with a monopoly on the import of blood and blood products into France, was necessarily at the centre of the affair.

The Lucas report spelled out the way in which approval for the blood tests for HIV developed in the United States in 1984 was deliberately held up so that the alternative under development at the Pasteur Institute could catch up; that was crucial at a time when the only means of screening blood donors was verbal enquiry into their lifestyle. When, later that year, it emerged that 6 per cent of blood donors in Paris were infected with HIV, Garretta calculated that the chance that all pools of blood collected in Paris were not contaminated must be less than one in a thousand.

Another element in the scandal was the delay in subjecting contaminated blood to the heat-treatment by then known to be effective in removing the risk of infection by hepatitis B. By the end of 1983, CNTS was negotiating with Immuno (of Austria) about the possibility of treating its existing stocks of blood, but delayed a decision (and responded only with delay to an offer by the blood products centre at Lille to collaborate on the heat-treatment of all blood used for transfusion and the manufacture of blood products in France).

During the summer of 1985, while Allain was conducting a controlled trial of the consequences of using heat-treated and untreated Factor VIII for administration to haemophiliacs, CNTS promulgated the rule that treated blood products (then being imported from the United States) should be reserved for haemophiliacs showing no signs of antibodies against HIV antigens, and that those which had already seroconverted (roughly half) should be given untreated Factor VIII "while stocks last". In May 1985, it was decided that the French health service would no longer meet the cost of untreated Factor VIII after 1 October.

Among the circumstances contributing to this delay identified by Lucas and the judgement of the first trial were Garretta's view of CNTS as an industrial enterprise founded on French research and the incestuous relationship between CNTS and the French Association of Haemophiliacs (AFH), representing physicians concerned with the treatment of haemophiliacs. The judgement of the first trial complains that AFH, whose secretariat and office space were provided by CNTS, was never told with sufficient clarity that all stocks of blood collected in France were probably contaminated with HIV.

The human tragedy is immense. Half of France's 6,000 haemophiliacs were infected with HIV, perhaps 1,200 of them during the period when they might have been protected. Why should the anger generated give way to leniency? One powerful consideration is that those on trial are only a subset of those involved in the decision. Schnerb likened them (or at least his client) to people who had reported a tragic road accident to the police, and had been arrested for their pains.

All those arraigned are on record during 1984 and 1985 as having warned others of the dangers of using contaminated blood. The doubt is whether they warned the right people, and with sufficient urgency. Schnerb last Wednesday, on behalf of Allain, had no doubt. His client was a "visionary", who among other things had written to Garretta in January 1985 to urge immediate steps to heat-treat the contaminated blood. "If you condemn him now", Schnerb insisted, nobody will ever again write such a courageous letter. But his most telling point remained the claim that there should have been a hundred defendants, not just four. It is the system that is on trial.

John Maddox

Comet on target for Jupiter

Clark R. Chapman

SIXTY-FIVE million years ago, a comet or asteroid exploded in Mexico with the force of perhaps a thousand million megatonnes of TNT. Evidence mounts that this once-in-a-hundred-million-year event dramatically affected the evolution of life on our planet. Who would have thought that we might ourselves witness — next year, indeed — a cosmic impact of comparable proportions? This time, Jupiter is the target.

A fascinating comet discovered earlier this year may, in fact, dive into Jupiter in

Jupiter's atmosphere, and new ideas about the nature of comets.

Planetary geologist Eugene Shoemaker (U.S. Geological Survey, Flagstaff, Arizona) told the story of the comet's discovery at the American Geophysical Union meeting on 25 May, in Baltimore, Maryland. He had just been presented with the union's Fred L. Whipple Award. On 25 March, Gene's wife, Carolyn, noted a "squashed comet" on a pair of photographs taken with the 18-inch Schmidt camera on Mount Palomar, California

actually orbiting Jupiter, in a loosely bound, nearly parabolic orbit . . . and that they are slowly spreading apart. From backward extrapolation, it is clear that the parent comet split apart just last summer. Furthermore, orbital calculations extrapolated backwards now demonstrate that Comet Shoemaker–Levy 9 passed only 100,000 km from Jupiter's centre-of-mass around July 1992. That means that it penetrated within Jupiter's Roche limit, where the giant planet's tidal forces are theoretically capable of pulling a weak, cohesionless body apart.

Comets have been observed to split before, although usually during close passage to the Sun. Never before has there been such a clear case of disruption of a

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Heading for disaster: Comet Shoemaker–Levy 9, photographed by D. Jewitt and J. Luu, seems destined for a spectacular demise next year

late July 1994. But for the happenstance that nature has apparently handed us a tails instead of a heads in a cosmic flip of the coin, the event could have been the most spectacular astronomical event ever to be witnessed in the heavens during recorded history. Unluckily, however, Comet Shoemaker–Levy 9 seems to be headed for Jupiter's far side, so the giant planet itself will block the celestial light show from our vantage point. We can only regret that Voyager's cameras have been permanently turned off and that Galileo will arrive 16 months too late to record the event.

Nevertheless, the planetary science community is buzzing with the news. While citizens in the street may not be able to witness the explosions directly, a wealth of phenomena should be observable through Earth-based telescopes. And the implications of Comet Shoemaker–Levy 9 for understanding our planetary system are only beginning to be appreciated. After this cosmic drama is played out, we may have new evidence about how planetary rings are formed and how satellites are cratered, observational data about multi-million-megatonne atmospheric explosions, a kind of pre-Galileo probe of

(*IAU Circ.* No. 5507). The photos were part of a search programme for comets and asteroids that the Shoemakers are conducting with the assistance of premier amateur comet discoverer, David Levy. (In late May, Carolyn Shoemaker discovered her thirtieth comet; she is closing in fast on the all-time record number of comets to be discovered by an individual astronomer.)

The Shoemaker team immediately contacted Jim Scotti, who was using the larger Spacewatch telescope on Kitt Peak in Arizona, to confirm their peculiar comet. He soon reported that the comet, situated near Jupiter in the sky, was resolved into a short string of several separate comets. (Later photos, see above, taken from the Mauna Kea Observatory show about 20 separate pieces, strung out like separate railroad cars along their orbital track around the Sun.) The ensemble of comets was designated 1993e (fifth comet discovered or recovered in 1993) and named Periodic Comet Shoemaker–Levy 9 by Brian Marsden of the International Astronomical Union's Central Bureau in Cambridge, Massachusetts.

Subsequent positional measurements soon revealed that the train of objects is

body by tidal splitting. Tidal splitting of a large object, near the Earth, had been suggested a couple of decades ago by George Wetherill (Carnegie Institution, Washington, DC) as a possible way of producing the unusual size distribution of objects that cratered the Moon and inner planets during the 'late heavy bombardment' about four billion years ago. But subsequent theoretical work had cast doubt on the idea that a large body would actually come apart during the few minutes or hours it spent within the Roche zone.

Was the new comet a large enough body to resolve this controversy, or was it just another small comet? (Some small comets split for no known reason at all, not necessarily implicating tidal forces.) Indeed, there is some question about whether Comet Shoemaker–Levy 9 is a normal comet at all. At the moment, it is actually a satellite of Jupiter. And preliminary observations reveal little evidence for the gaseous emissions typical of new arrivals from the distant Oort cloud of comets. (Although it is too cold near Jupiter for water molecules to be liberated, evidence for much more volatile ices might have been expected from a freshly

disrupted comet.) Perhaps this 'comet' is more akin to a Trojan asteroid, locked on Jupiter's solar orbit, or one of Jupiter's outer satellites, in composition. But then nobody really knows what Trojan asteroids are made of, either.

The visible haze around the train of comets is mostly dust, illuminated by reflected sunlight, and the cores of the separate pearls on the string may be nearly bare fragments instead of distended cometary atmospheres veiling much smaller nuclei. On that assumption, Gene Shoemaker estimates that the largest fragment could be 10 km across. The original object before break-up might have been fully 20 km in diameter. That is the approximate size of the object responsible for the 200-km diameter Chicxulub crater in Mexico, formed 65 million years ago.

Shoemaker spoke at an Asteroid Hazards conference, held in Erice, Sicily, in late April about his view that the Chicxulub impactor was not alone in helping to create the dramatic environmental consequences that shaped the sudden transition (K/T boundary) from Cretaceous to Tertiary biota. Showing a dramatic picture of the then recently discovered Comet Shoemaker-Levy 9, he advocated cometary split-up and a 'comet shower' as the fundamental mechanism for the ecological holocaust. In the Erice audience was the American physicist Edward Teller, father of the hydrogen bomb. Teller's reaction to the week-long discussions about cosmic threats to life on Earth was, mysteriously, to express a profound curiosity about Jupiter's atmosphere. In particular, Teller queried University of Arizona cosmochemist John Lewis about the Red Spot, its nature and its origin.

That was several weeks before Brian Marsden announced on 22 May, in *IAU Circ.* No. 5800, that orbital calculations of Comet Shoemaker-Levy 9 show that it is aimed to pass well inside Jupiter's radius of about 75,000 km: the comet seems destined to collide with the planet. Don Yeomans and Paul Chodas, of the Jet Propulsion Laboratory, have made independent calculations that agree with Marsden's. Their latest results (1 June) indicate an approach to within about 35,000 km of Jupiter's centre of mass around 20 July 1994. Unlike their widely reported differences last year concerning the future approach of the Comet Swift-Tuttle to Earth (see *Nature* **360**, 623; 1992), Yeomans and Marsden agree this time that there is a much better than fifty-fifty chance that at least some of the comet will collide with Jupiter next year.

According to Andrea Carusi (Istituto di Astrofisica Spaziale, Rome), it is very likely that either the entire comet train will fall into Jupiter (over the course of several days) or else all the objects will miss. Yeomans says that the impacts will

nominally occur in Jupiter's southern mid-latitudes on the nightside, perhaps 20° or 30° around the planet's limb from our vantage point. (All of these numbers are highly uncertain.)

The impact could certainly rival or exceed the K/T holocaust itself in terms of explosive energy. Crashing into Jupiter at 60 km s⁻¹, the largest fragment would liberate a billion megatonnes of energy. John Lewis estimates that a fireball formed perhaps 1,000 km below Jupiter's cloud deck might ascend through the clouds in an awesome display. If the comet fragments were to hit the front side of Jupiter, the explosions would be visible from Earth in broad daylight as the separate objects followed each other down into Jupiter's clouds.

Given that the impacts are expected to occur on the nightside, astronomers will be observing less dramatic phenomena for evidence of the comet's demise. Reflected light from Jupiter's moons might enable them to measure characteristics of the explosions, provided they aren't snuffed out from visibility in the depths of Jupiter's atmosphere. And, a few hours after each impact, the entry point will rotate into view from Earth. Will a new 'red spot' form, validating Edward Teller's premonitions, as a result of an explosion exceeding the greatest superbomb he had ever imagined, or will the comet fragments be swallowed up by Jupiter with nary a trace?

If the comet, as now seems less likely, misses Jupiter, will we be able to watch it begin to form a new ring around Jupiter, following the tidal break-up sequence

proposed two years ago by Luke Dones (now at NASA Ames Research Center)? Dones suggested in *Icarus* (**92**, 194–203; 1991) that Saturn's ring was formed by tidal break-up of a Chiron-like giant comet, which he hypothesized to have passed within the Roche zone of Saturn during the past hundred million years. Dones thinks that the so-called population 2 craters on Saturn's inner moons might have been created by other debris from the same break-up. Might a few fragments from the recent break-up of Comet Shoemaker-Levy 9 start raining down on one of Jupiter's moons, forming a few new impact craters for the Galileo camera to observe before the spacecraft's Jupiter mission ends in the late 1990s?

These are just some of the interesting questions being discussed on the electronic mail circuit by planetary scientists during the first week following release of Brian Marsden's *IAU Circ.* No. 5800. If past history is any indication, some of the supposed facts will be found to be wrong. And many more implications of this remarkable turn of events will be thought of. Marsden guesses that by the end of Jupiter's current apparition two months from now, the probabilities about the impending impact and its geometry will be converted into certainties. Then astronomers will begin to organize the world-wide Jupiter/Comet Watch to coordinate studies of the probable demise of what already is surely the most interesting comet ever discovered. □

Clark R. Chapman is at the Planetary Science Institute, Tucson, Arizona 85705, USA.

CATALYSIS

Starting from first principles

Joachim Sauer

THE design of efficient and selective catalysts is still largely empirical, despite the substantial efforts made to elucidate the mechanisms of heterogeneous catalysis. For acidic zeolites Kramer *et al.* now show, on page 529 of this issue¹, that reliable reactivity predictions can be made by combining the power of quantum chemical techniques, starting from first principles, with classical methods of lattice-energy minimization. Their results neatly explain why two catalysts have different reactivities although they have the same composition and the same number and type of active sites, and differ only in the structures of their aluminosilicate frameworks.

Acidic zeolites can be viewed as three-dimensional networks of corner-sharing SiO₄ tetrahedra, the spaces between these forming molecular-scale micropores of varying shape and dimension. A few sili-

con ions, about one in 10 to one in 50, are replaced by aluminium ions, and protons are added locally for charge compensation. The protons are localized at one of the four oxygen atoms of the AlO₄ tetrahedron forming 'bridging hydroxyl' groups. These sites are known to be active in catalytic conversion of hydrocarbons².

It is a great advantage, both in the exploitation of zeolites and in modelling their properties, that these hydroxyl groups are part of the periodic structure and are homogeneously distributed over the 'internal surface' of the crystal. The average distance between them is large enough for interactions to be negligible, but their concentration is high enough to give high conversion rates per gram of catalyst. That the catalytic conversion takes place within micropores of molecular dimensions adds the appealing feature of shape selectivity to the acidic function

of zeolite catalysts.

These properties, amongst others, explain the increasing popularity of zeolites for industrial acid-catalysed hydrocarbon conversion, both in oil refining and petrochemical processes. For example, zeolites replace environmentally hazardous acidic aluminium chloride in benzene alkylation processes (the Mobil Badger process). With the increasing use of natural gas as a feedstock in the manufacture of petrochemicals, there is keen interest in the methanol-to-petrol and methanol-to-alkene reactions, in which C-C bond coupling is a key step.

Kramer and colleagues¹ have made a leap towards the ambitious goal of understanding the mechanism of such complex reactions, even if their study is of the simple exchange reaction of deuterium for hydrogen in methane. In itself this reaction is of little interest, but the transition complex formed, a carbonium ion in close contact with an AlO_4^- surface site, may well be an intermediate in hydrocarbon cracking reactions³, in which lighter alkanes, alkenes and molecular hydrogen may be formed. Their results are not, of

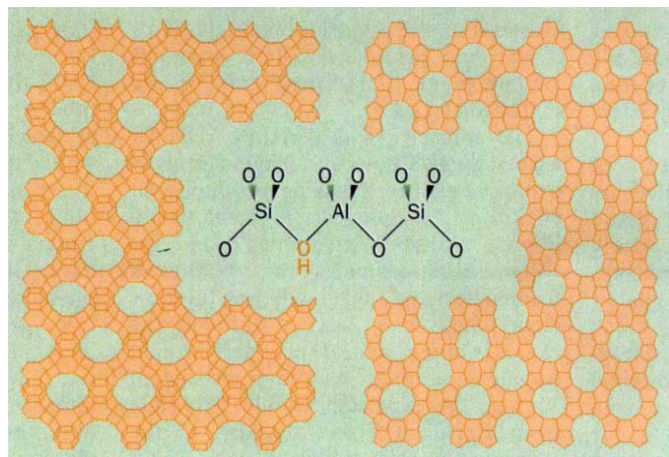


FIG. 1 The acidity of zeolite catalysts originates from bridging hydroxyl groups (inset). Faujasite (left) and ZSM-5 (right), different crystal forms of the same compound, have different activities for deuterium-hydrogen exchange, although the number of active sites is about the same. Why this is so is revealed in the theoretical study of Kramer *et al.*¹.

course, the definitive answer to the question of whether carbonium ions exist as intermediates on the surface of zeolite catalysts. More complete searches of the potential energy surface will come later, drawing on larger models of the active site (which are more realistic, but require more computer time). Moving on to C_2H_5^+ or larger carbonium ions, rather than using methane, will allow the study of hydrogen elimination and C-C bond cleavage reactions.

For intermolecular reactions in the gas phase, quantum chemical *ab initio* techniques have successfully predicted the structures, vibrational frequencies and relative energies of the reactant and the transition species⁴. These are the ingredients needed to calculate rate constants by means of transition-state theory.

In heterogeneous catalysis, the application of quantum chemical techniques is complicated because, on the molecular scale, the dimensions of a solid catalyst are essentially infinite. Exploiting translational symmetry, as physicists do in band-structure calculations, is not much help because of the random distribution of active sites and the enormous size of zeolite unit cells (hundreds of atoms).

One way to overcome these difficulties is to confine the quantum chemical treatment to the atoms of the active site and their main interactions. Molecular models such as that adopted by Kramer *et al.* have had remarkable success in predicting the local structure of bridging hydroxyl groups, their spectroscopic parameters and their interaction with external molecules. But this 'molecular' approach has a serious drawback: the results refer to one isolated active site, and do not account for effects due to the structure of the aluminosilicate framework.

Where Kramer *et al.* break away from

previous calculations on simple reactions at bridging hydroxyl groups⁵⁻⁷ is in combining *ab initio* results for a single active site with theoretical predictions of how the reactivity changes when the sites are part of a given framework. How did they achieve this? Periodic *ab initio* calculations on zeolites, as I stressed earlier, are not feasible because of the size of the unit cells. But lattice-energy minimizations using classical ion-pair or molecular-mechanics potentials are computationally much less demanding. They can be used to determine the relative stabilities of zeolite structures with the aluminium ions in crystallographically different positions and with the proton attached to crystallographically different oxygen

atoms. The problem is, of course, the reliability of the potentials used, but the increasing body of *ab initio* data becoming available for finite models of zeolites is setting these on firmer ground.

Kramer *et al.* use the results of one such technique, which predict that the average difference in proton affinity between neighbouring oxygen atoms of the active site is higher if their catalyst has the faujasite structure (FAU) than if it has the ZSM-5 structure (MFI). From their *ab initio* calculations on the molecular model catalyst they know that the reaction barrier height increases with increasing proton affinity difference between the two oxygen atoms involved in the deuterium-hydrogen exchange. These two pieces of information together explain why the exchange rate is lower on the catalyst FAU than on the catalyst MFI.

This result is remarkable for a further reason. Until now it was generally believed that the activity of acidic zeolite catalysts depended on the proton affinity of the bridging hydroxyl oxygen atom. Kramer and colleagues are now telling us that, for this reaction at least, the key quantity is the difference in proton affinity between this oxygen atom and the next. □

Joachim Sauer is in the Quantum Chemistry Group of the Max Planck Society at Humboldt University, O-1199 Berlin, Germany.

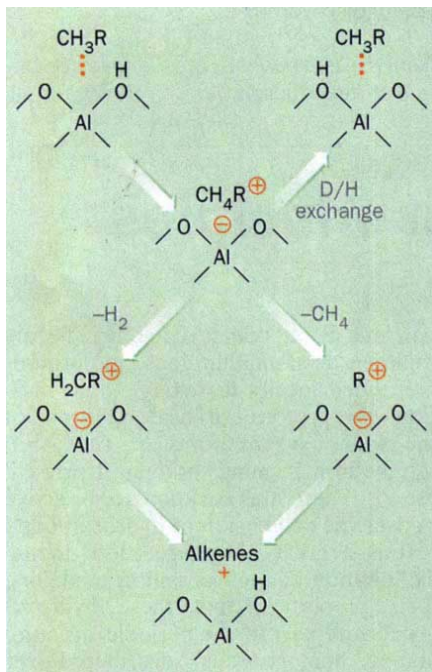


FIG. 2 Possible role of carbonium ion surface species in hydrocarbon conversions on zeolite catalysts. Whether these species are intermediates (as frequently postulated) or transition structures (as Kramer *et al.* predict for CH_5^+) is not clear. Top: hydrogen-hydrogen exchange mechanism (seen as D/H exchange in the new experiments); bottom: H_2 elimination and C-C bond cleavage.

1. Kramer, G. J., van Santen, R. A., Emeis, C. A. & Nowak, A. K. *Nature* **363**, 529-531 (1993).
2. Haag, W. O., Lago, R. M. & Weisz, P. B. *Nature* **309**, 589-591 (1984).
3. Haag, W. O. & Dessau, R. M. *8th Int. Congr. Catalysis*, 305-316 (Dechema, Berlin, 1984).
4. Hehre, W. J., Radom, L., Schleyer, P. v. R. & Pople, J. A. *Ab Initio Molecular Orbital Theory* (Wiley, New York, 1986).
5. Sauer, J., Kölmel, C. M., Hill, J.-R. & Ahlrichs, R. *Chem. Phys. Lett.* **164**, 193-198 (1989).
6. Sauer, J., Horn, H., Häser, M. & Ahlrichs, R. *Chem. Phys. Lett.* **173**, 26-32 (1990).
7. Kazansky, V. B. & Senchenya, I. N. *J. molec. Catal.* **74**, 257-266 (1992).

A gene for neurofibromatosis 2

Kenneth W. Kinzler and Bert Vogelstein

A MILESTONE in cancer genetics was marked 26 years ago by the discovery of chromosome 22 monosomy in meningiomas, the first specific chromosomal change identified in solid tumours¹. A sense of closure for this piece of history has now been achieved with the isolation of the gene responsible for the type 2 form of neurofibromatosis^{2,3}.

The neurofibromatoses are character-

identification of subtle intragenic mutations in the germline of NF2 patients. Rouleau *et al.* found 16 such mutations, 15 of which were predicted to result in truncated proteins.

These exciting findings lay the groundwork for the direct presymptomatic diagnosis of NF2, a condition that affects about 1 in 37,000 individuals. Moreover, the observation by Rouleau *et al.* of

theory of tumour suppressor genes suggests that mutation or loss of the normal allele is the rate-limiting event¹². It is notable that six of eight tumours containing mutations of NF2 exhibited loss of the wild-type allele^{2,3}. Although this result implies that mutation of both alleles of NF2 provides a growth advantage, it does not prove that both are required for tumour initiation.

As the table shows, the products of suppressor genes operate throughout the cell and affect a variety of regulatory processes. Sequence analysis of the predicted NF2 gene product (designated

Dominantly inherited genetic predispositions to tumours

Syndrome	Predominant tumour types	Chromosome	Gene	Location/function
Early-onset familial breast cancer	Breast carcinoma	17q	?	?/?
Familial adenomatous polyposis	Multiple colorectal adenomas	5q	APC	Cytoplasmic/?
Familial melanoma	Cutaneous malignant melanoma	9p	?	?/?
Gorlin syndrome	Basal cell carcinomas of the skin	9q	?	?/?
Hereditary non-polyposis colorectal cancer	Colon and uterine carcinomas	2p	?	?/Replication fidelity?
Li-Fraumeni	Leukaemia, soft-tissue sarcoma, osteosarcoma, brain tumour, breast and adrenal cortical carcinomas	17p	p53	Nuclear/transcriptional regulation
Multiple endocrine neoplasia type 1	Parathyroid, endocrine pancreas and pituitary tumours	11q	?	?/?
Multiple endocrine neoplasia type 2	Medullary thyroid carcinoma, pheochromocytoma	10	?	?/?
Neurofibromatosis type 1	Multiple peripheral neurofibromas	17q	NF1	Cytoplasmic/GTPase-activating protein
Neurofibromatosis type 2	Central schwannomas and meningiomas	22q	NF2	Cytoskeletal-membrane link
Retinoblastoma	Retinoblastomas, osteosarcomas	13q	Rb	Nuclear/transcriptional regulation
von Hippel-Lindau	Renal-cell carcinoma	3p	?	?/?
Wilms tumour	Nephroblastoma	11p	WT1	Nuclear/transcriptional regulation

ized by a predisposition to tumours of the nervous system. In neurofibromatosis type 2 (NF2), this predisposition results in schwannomas and meningiomas of the cranial nerves and the spinal nerve roots. In contrast, the more common neurofibromatosis type 1 (NF1) is characterized by peripheral neurofibromas. These diseases were shown to be genetically distinct when the genes responsible were mapped to separate chromosomes in 1987 (refs 4–6). Three years later, the NF1 gene was cloned from chromosome 17 (refs 7–9).

Now, in separate studies, one in *Cell*² and one on page 515 of this issue³, Trofatter *et al.* and Rouleau *et al.* describe the identification of the NF2 gene from chromosome 22. In both cases, the investigation centred on NF2 patients with gross deletions in the critical region. Through elegant application of positional cloning, the region containing the deletions was isolated and a suspect gene identified. Irrefutable evidence for the conviction of this suspect came from the

somatic mutations in sporadic meningiomas and schwannomas strongly suggests that NF2 is the target of the chromosome 22 monosomies. Chromosome 22 losses have also been reported in tumours outside the nervous system, including those of the colon¹⁰ and breast¹¹, and it will be interesting to see what involvement, if any, NF2 has there.

NF2 now joins a number of dominantly inherited, tumour-predisposition genes that have been cloned or localized (see table). Basic questions about the biology of all of these genes remain unanswered. For example, what determines the tissue distribution of tumours in these diseases? Many of the genes (including NF1 and NF2) are expressed and presumably required in a wide variety of tissues. Why should tumours in neurofibromatosis kindreds be limited to the nervous system? And what determines which neural cells will go on to form a tumour? All cells contain the inherited mutation, but only a minute fraction form a tumour. A classic

merlin or schwannomin) suggests that it occupies a novel niche. Although the first eight amino acids of merlin/schwannomin are unclear because of a single nucleotide difference between the reported sequences, both groups describe a protein of some 590 residues which is strikingly similar in sequence to moesin, ezrin and radixin, all members of the erythrocyte band 4.1 family. Proteins of this family are thought to link integral membrane pro-

1. Zang, K. D. & Singer, H. *Nature* **216**, 84–85 (1967).
2. Trofatter, J. A. *et al.* *Cell* **72**, 791–800 (1993).
3. Rouleau, G. A. *et al.* *Nature* **363**, 515–521 (1993).
4. Barker, D. *et al.* *Science* **236**, 1100–1102 (1987).
5. Seizinger, B. R. *et al.* *Cell* **49**, 584–594 (1987).
6. Rouleau, G. A. *et al.* *Nature* **329**, 246–248 (1987).
7. Wallace, M. R. *et al.* *Science* **249**, 181–186 (1990).
8. Viskochil, D. *et al.* *Cell* **62**, 187–192 (1990).
9. Cawthon, R. M. *et al.* *Cell* **62**, 193–201 (1990).
10. Okamoto, M. *et al.* *Nature* **331**, 273–277 (1988).
11. Chen, L. C. *et al.* *J. natn. Cancer Inst.* **84**, 506–510 (1992).
12. Knudson, A. G. *Cancer Res.* **45**, 1437–1443 (1985).
13. Bretscher, A. J. *Cell Biol.* **108**, 921–930 (1989).
14. Krieg, J. & Hunter, T. *J. biol. Chem.* **267**, 19258–19265 (1992).
15. Fazioli, F., Wong, W. T., Ullrich, S. J., Sakaguchi, K. & Appella, E. *Oncogene* **8**, 1335–1345 (1993).

teins with the cytoskeleton. This similarity is all the more intriguing given the observation that moesin, ezrin and radixin are tyrosine phosphorylated and may mediate growth-factor-specific cellular changes¹³⁻¹⁵.

Over the past five years, many (if not most) of the genetic loci responsible for familial cancers have been identified. The genetics have far out-paced understanding of the biochemistry and cellular biology of these genes' products. Some clues, such as

the predicted relationship of merlin/schwannomin to cytoskeletal elements, have emerged. But the clear challenge is now to unravel the mysterious process through which a defect in such a gene leads to a mass of tumour cells. □

Kenneth W. Kinzler and Bert Vogelstein are in the Molecular Genetics Laboratory, Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231, USA.

QUANTUM MECHANICS

Making waves with electrons

A. Zettl

A CENTRAL concept of modern physics is the wave-particle duality of matter. The wave-like nature of freely propagating particles can be inferred from various diffraction and scattering experiments. These methods typically extract the quantum-mechanical properties of matter from changes in the momentum or energy of interacting particles. On page 524 of this issue¹, Crommie *et al.* describe an experiment in which the wave nature of electrons is directly observed from standing-wave patterns spatially resolved on the surface of a clean copper crystal. They used a scanning tunnelling microscope to image the local density of states of electrons trapped in a two-dimensional layer at the crystal surface. Such visualization of purely quantum-mechanical interference phenomena brings gratifying reality to these typically ethereal concepts. One can imagine a variety of experiments where this ability could be used to extract new and useful information.

Scanning tunnelling microscopes (STMs) have been used in the past to study distinctive electronic behaviour ranging from charge density waves² to superconducting vortices³. Now we find

that the STM can also be used to image the quantum interference patterns of electrons moving in a two-dimensional plane. Although so-called two-dimensional electron gases are more usually associated with layered inorganic crystals, organic charge-transfer salts and artificial layered semiconductor structures, it turns out that electrons can be trapped at the close-packed surfaces of the noble metals, copper, silver and gold. Thus trapped, the electrons constitute a nearly ideal two-dimensional electron gas. Confinement to the surface has its origin in naturally occurring energetic barriers in directions perpendicular to the surface. On the vacuum side, electrons are trapped by the work-function (ionization) barrier; and on the bulk side, electrons see a forbidden band gap in the metal.

The resulting two-dimensional noble-metal surface states, which have been studied extensively in photoemission experiments^{4,5}, are to be distinguished from more localized 'Tamm' states, or dangling bonds, seen on semiconductor surfaces. Surface states have a long history of being studied with STMs. In fact, one of the first spectroscopic studies of metals with the STM was performed on gold, and clearly showed a spectroscopic signal due to a gold surface state⁶.

The ability to resolve the quantum-mechanical interference patterns of surface-state electrons provides

a new wrinkle to an old problem, and yields a new tool for the study of defects on metallic surfaces. Without defects, the density of the surface-state electrons would be flat and featureless; this is in contrast to charge density waves which exist independently of defects (see Figs 1 and 2). Structure in the electronic surface-state density arises from the scattering of the surface-state electrons from imperfections on the surface. This scattering causes the incident and reflected electron waves to interfere quantum mechanically and set up standing waves in the defects' vicinity. The physics behind this is similar to the Friedel oscillation⁷, a variation in total charge density around a defect (contributed to by electrons of all energies up to a maximum, the Fermi energy, which thus fixes the wavelength of Friedel oscillations).

An STM, though, does not measure the total density of electrons, but rather the local density of states at a given energy,

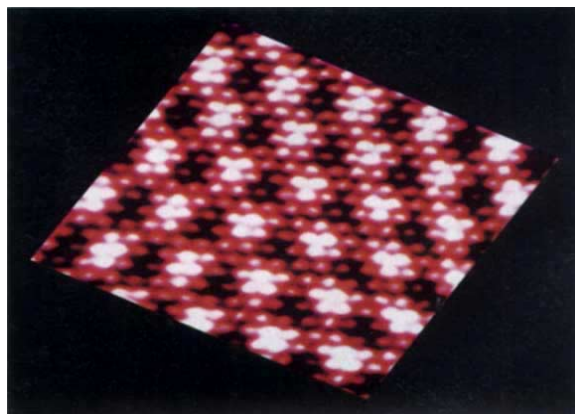


FIG. 2 Permanent wave. Charge density waves in TaS₂, revealed by atomic force microscopy. These are permanent corrugations of charge density, arising through a periodic lattice distortion driven by the electron-phonon interaction. (Courtesy of R. V. Coleman.)

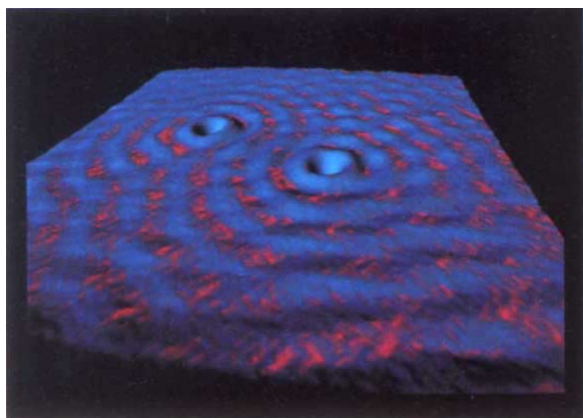


FIG. 1 Visibly ruffled. Standing waves (wavelength, 15 Å) of electrons scattered by a pair of point defects on the surface of a copper crystal. It is only through the interference effects created by the fixed defects that the electron waves become apparent. (Courtesy of D. M. Eigler.)

selected by choice of the voltage applied to the STM tip⁸. The STM, therefore, has the ability to resolve a Friedel oscillation by spatially mapping out the scattered electron density at different energies. The details of these oscillations are not solely dictated by the lattice periodicity but depend sensitively on the energy being probed, the electron's effective mass, and the scattering potential of the defect. By measuring an energy-resolved set of spatial oscillations one can work backwards and extract information about the defect scattering potential, as well as surface-state electronic parameters. It is of course well appreciated that defects, such as step edges, are important in regulating such processes as film growth and catalysis on metal surfaces.

The experiment reported by Crommie *et al.* used a low temperature (4 K) ultrahigh vacuum STM. Recently Hasegawa and Avouris of IBM Yorktown

Heights have seen similar oscillations near step edges on a gold surface using a room-temperature STM (unpublished results). Although there are some differences, the basic mechanism of quantum interference of surface-state electrons appears to operate on both the copper and gold close-packed surfaces. A careful study of differences between the room-temperature and 4 K results could possibly lead to a better understanding of inelastic scattering of surface-state electrons.

Perhaps the most interesting possibility raised by these results is the potential for studying the interaction of surface-state electrons with adsorbates, such as gas molecules attached to catalyst surfaces. The interest here is twofold. First, the interference patterns of the surface-state electrons might be used as a probe of the electronic structure of an individual adsorbate. Second, if the adsorbates do scatter the surface-state electrons, then one might be able to arrange the adsorbates with the STM tip⁹ to create microscopic

electron containers of arbitrary shape. Confined electron structures in buried semiconductor interfaces have been studied for years using transport and optical probes. The possibility now exists directly to access and spatially resolve the state density of artificially confined two-dimensional electrons, ideally including two-dimensional electrons in the presence of strong magnetic fields. □

A. Zettl is in the Department of Physics, University of California, Berkeley, California 94720, USA.

1. Crommie, M. F., Lutz, C. P. & Eigler, D. M. *Nature* **363**, 524–527 (1993).
2. Coleman, R. V., Drake, B., Hansma, P. K. & Slough, G. *Phys. Rev. Lett.* **55**, 394–397 (1985).
3. Hess, H. F. *et al. Phys. Rev. Lett.* **62**, 214–217 (1989).
4. Gartland, P. O. & Slagsvold, B. J. *Phys. Rev. B* **12**, 4047–4058 (1975).
5. Kevan, S. D. *Phys. Rev. Lett.* **50**, 526–529 (1983).
6. Kaiser, W. J. & Jaklevic, R. C. *IBM J. Res. Develop.* **30**, 411–416 (1986).
7. Friedel, J. *Nuovo Cim. Suppl.* **7**, 287–311 (1958).
8. Tersoff, J. & Hamann, D. R. *Phys. Rev. B* **31**, 805 (1985).
9. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524–526 (1990).

CELL PHYSIOLOGY

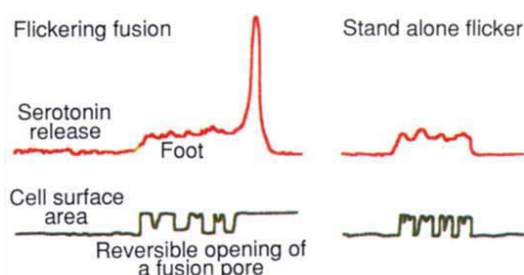
Secretion without full fusion

Erwin Neher

It seems terribly wasteful that, during the release of hormones and neurotransmitters from a cell, the membrane of a vesicle should merge with the plasma membrane to be retrieved for recycling only seconds or minutes later. If parsimony has any meaning in biology, one might expect nature to have designed a way of releasing neurotransmitters and hormones which avoids intermixing components of the vesicular membrane with those of the plasma membrane. A study by Alvarez de Toledo *et al.* on page 554 of this issue¹ now points to a way that reconciles efficiency with classical concepts of vesicular exocytosis.

Several years ago^{2,3}, investigations into electrical membrane capacitance indicated that fusion of large secretory granules with the plasma membrane may be a transient phenomenon. Because the plasma membrane is, electrically speaking, a capacitor that is proportional to surface area, a change in capacitance can be observed when vesicular membrane is inserted into the surface membrane during exocytosis. For large granules, like those of mast cells, step-like increases can indeed be resolved when single granules fuse. Such records show that not all fusion events are permanent, but that the extra membrane often appears and disappears several times in the electrical measurement before the process is finalized. This phenomenon was termed 'flickering fusion'. Breck-

enridge and Almers³, and Zimmerberg *et al.*⁴, studying details of complex admittance changes in giant granules of beige mouse mast cells, concluded that a fusion pore forms as an initial electrical pathway between the interior of the granule and the extracellular space. This fusion pore gradually enlarges for complete fusion or — in the case of flickering fusion — opens



Measurement of cell surface area (lower traces) of mast cells undergoing degranulation shows that exocytosis can be reversible in the sense that the contribution of a granule to the surface membrane appears and disappears repeatedly. Simultaneous measurement of released substance (upper traces) indicates that serotonin seeps out of the granule at a low rate. The bulk of the granule contents is released after the increase in surface area becomes permanent. Occasionally 'stand alone' flickers are observed, during which exocytosis does not run to completion. Then, only a small fraction of granule contents is released from large vesicles. Comparison of kinetics and amounts of release between granules of different sizes indicates that small synaptic vesicles may be able to shed most of their contents in less than a millisecond through a reversible fusion pore.

and closes reversibly. Studying the fluorescence of quinacrine, loaded into the granules, it was concluded that release of the fluorophore was negligible during the flickering phase of secretion and that true release required irreversible fusion³.

This conclusion, which previously was thought to be valid for large granules, may now have to be reconsidered, particularly for the case of smaller ones. Measurements with a more sensitive detection technique indicate that smaller granules may release much of their contents by diffusion through a reversible fusion pore. Extrapolating the available data to the size of synaptic vesicles, it seems possible that these small structures shed their contents before stable fusion has occurred, within a fraction of a millisecond.

The detection technique concerned is amperometry, the electrochemical detection of released substances by carbon-fibre microelectrodes. This technique, in its micro version, was pioneered by Wightman *et al.*⁵, who showed that spike-like currents could be detected when the tip of a carbon-fibre microelectrode was located near a stimulated adrenal chromaffin cell. All the properties of these current transients were compatible with the idea that they represent the release of the contents of single catecholamine-containing granules. Chow and others⁶ (including myself) corroborated this interpretation and discovered some fine structure to these secretory events. Conspicuously often, the main transient portion of the signal is preceded by a pedestal (or 'foot'). This foot is not compatible with what one would expect on the basis of diffusion theory, if release occurs instantaneously at some distance from the detecting surface. Rather, we concluded that the foot represents a trickling release through a fusion pore before bulk exocytosis occurs.

Alvarez de Toledo *et al.* now demonstrate that this is indeed the case by combining the amperometric technique with capacitance measurements. They worked on giant granules of beige mouse mast cells, where the conductance of the flickering fusion pore can be measured. In these granules, amperometry detects the neurotransmitter serotonin⁷, which is contained in the granules together with histamine. The authors observe an amperometric foot when, and only when, there is flickering fusion. They also occasionally observe 'stand alone' feet when there are capacitance flickers that do not result in a permanent increase in capacitance (see figure).

One notable finding is that plots of amperometric signal amplitude against fusion pore conductance for different granules have different slopes, showing that the concentration

of free mediator inside the granules differs between granules. Estimates of the total content within granules further indicate that the total concentration of mediator is much higher than that of free mediator alone.

Another interesting aspect is the comparison of foot parameters for granules of different sizes. The giant granules of beige mast cells have mean diameters around 2.5 μm . Normal mast cell granules, on the other hand, have diameters of around 0.7 μm , which implies a 50 times smaller volume, and, most likely, 50 times less substance to secrete. If the gateway through which release occurs opens up in a similar fashion in the two cases, secretion should occur 50 times faster for the smaller granules. Likewise, the proportion of contents being released during flickers or during a foot should be larger.

Alvarez de Toledo *et al.* compared events in normal and beige mast cell granules. They find the time course of secretion is much quicker in normal granules, although not by a factor of 50; they also find that the foot carries a larger proportion of the total amperometric signal, although it is shorter with the smaller granules. They compare their values (foot-length, duration of the main signal and proportion of contents secreted during the foot) with those recently measured for adrenal chromaffin cells³ (granule diameter is about 0.25 μm) and find these trends to be continued. When plotted in double-logarithmic fashion against vesicle diameter, the values of all three quantities are found to lie on straight lines.

Extrapolated to a diameter of small synaptic vesicles, the prediction is that the foot (or the flicker) should be submillisecond in duration and should release around 8 per cent of the vesicle contents. If, unlike dense core granules, the contents were freely diffusible, release should be even faster and a vesicle might well shed most of its contents during a flicker. The vesicle membrane might be recycled without ever undergoing full fusion. Admittedly, this is speculation — but, if true, it might provide a reason why vesicles in fast synapses are small. $\square$

Erwin Neher is in the Department of Membrane Biophysics, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg, D-3400 Göttingen, Germany.

METEORITE ORIGINS

Blasting rocks off planets

H. J. Melosh

CAN rocks from the surface of a major planet or satellite be launched into interplanetary space by natural processes? A few years ago the answer to this question would be a resounding “no” from the experts on both volcanism and impact cratering, the only geological processes known to eject solid material at substantial velocities. Observation, however, has once again confounded expectation. Research in the past decade has established beyond reasonable question a lunar origin for some meteorites¹. Furthermore, a strong case has been made that the SNC (Shergottite, Nakhilite and Chassigny) meteorites originated on Mars². Although

ejection was accompanied by no more than 200 kbar of shock¹. Of the martian meteorites, some of the Shergottites were shocked to pressures as high as 300–450 kbar², but Nakhala records only very mild shock, 50–100 kbar⁴, more than an order of magnitude less than that required for escape by the pressure–velocity relation.

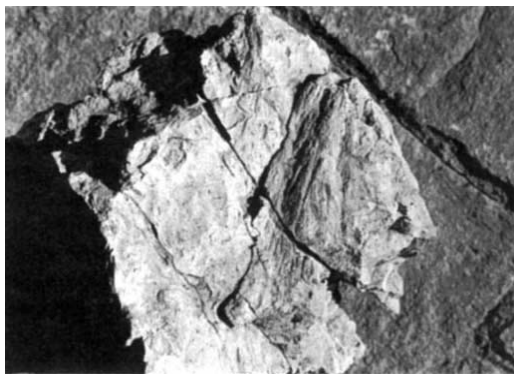
The problem with the pressure–velocity relation is that it applies only to material completely engulfed by the shock wave. Very close to the target surface, however, the ambient pressure is zero. No matter how strong the impinging shock wave, the free surface can never be raised to a pressure higher than zero: this effectively

shields surface rocks from strong compression. A strong pressure gradient develops below the surface, however, as the pressure rises rapidly from zero at the surface itself to the full bulk-shock pressure a short distance below. This gradient translates into powerful acceleration by Newton's first law, and the net result is high-speed ejection of lightly shocked near-surface material^{5,6}.

In the experiment performed by Gratz, Nellis and Hinsey, a coin-shaped disk of aluminium 19 mm in diameter and 3 mm thick was fired face-on into a slab of Westerly granite at 4 km s⁻¹. The rock directly beneath the impact was crushed and shocked to the expected 400 kbar. The major innovation in this experiment, however, was to place a cylindrical mass of low-density foam in front of the target, thereby catching any

high-speed ejecta thrown back off the surface. This arrangement worked magnificently: the powerful spray of fast ejecta actually ripped the foam cylinder apart along a 40° cone. An ejection velocity of about 1 km s⁻¹ was indicated by the distance of penetration of ejecta particles into the foam. In spite of this high ejection speed the recovered particles showed little or no sign of shock, mostly less than 50 kbars, and fragment sizes ranged up to several millimetres.

This experiment thus verifies the theoretical prediction of a small quantity of fast, lightly shocked ejecta from the near-surface zone of the target, and explains how rocks from a planet as large as Mars may be ejected into interplanetary space by large impacts. Even ordinary meteorites may require a mechanism of this type to spall them from their parent asteroids and throw them into orbits from which resonances with Jupiter's orbit may eventually perturb them into Earth-



This 22-cm block of Malm limestone was ejected from the 24-km diameter Ries impact crater in southern Germany about 15 million years ago and landed nearly 200 km away, near the present town of St Gallen, Switzerland. The prominent shatter cone indicates that it was shocked to about 10 kbar, far less than the 220 kbar shock pressure normally associated with its ejection velocity of 1.4 km s⁻¹. (Courtesy of B. A. Hofmann.)

volcanic ejection still seems incapable of achieving planetary escape velocity, the ejecta from large impacts are not so limited. Small-scale experiments by Gratz, Nellis and Hinsey, reported on page 522 of this issue³, show that impacts can eject a small quantity of lightly shocked material at high velocity.

Older work on the maximum velocities achieved by impact ejecta focused on the relation between the pressure in the shock wave generated by the impact and the velocity of material just behind the shock wave. Measured directly in laboratory experiments, this relation indicated that to accelerate surface rocks to planetary escape velocity (2.4 km s⁻¹ for lunar ejecta and 5.0 km s⁻¹ for martian ejecta), pressures of roughly a megabar must develop (440 kbar and 1.5 Mbar for lunar and martian basalts, respectively). Such high pressures pulverize or melt the ejected material. However, petrographic study of the lunar meteorites indicates that their

1. Alvarez de Toledo, G., Fernández-Chacón, R. & Fernández, J. M. *Nature* **363**, 554–558 (1993).
2. Fernández, J. M., Neher, E. & Gomperts, B. D. *Nature* **312**, 453–455 (1984).
3. Breckenridge, L. & Almers, W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1945–1949 (1987).
4. Zimmerberg, J., Curran, M., Cohen, F. S. & Brodwick, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1585–1589 (1987).
5. Wightman, R. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 10754–10758 (1991).
6. Chow, R. H., von Rüden, L. & Neher, E. *Nature* **356**, 60–63 (1992).
7. Tatham, P. E. R., Duchon, M. & Millar, J. *Pflügers Arch.* **419**, 409–414 (1991).

crossing orbits. The recent discovery⁷ of a number of 5–10-km diameter asteroids in the vicinity of Vesta, which are most plausibly fragments ejected from the crust of that body, also points to the importance of near-surface interactions accompanying a large impact (see McKinnon's News and Views article⁸). On Earth, 20-cm fragments of Malm limestone, the uppermost rock unit at the site of the Ries impact crater in Southern Germany, have been found in Switzerland, nearly 200 km from the impact site^{9,10} (see figure). These blocks again show a combination of low shock damage and high ejection velocity (1.4 km s^{-1}), although in this case the velocity was not high enough to throw them off the planet.

The experiment of Gratz *et al.* is not, of course, the last word on the subject. A more quantitative comparison between the experimental results and theory is very important. Theory will then be needed to extrapolate from the laboratory results to

the higher velocities and much larger projectile sizes involved in Solar System impacts. Nevertheless, this simple experiment has given a major boost to ideas on how rocks may be lofted from planets into interplanetary space. It should also put meteoriticists on the lookout for meteorites from other large planets — Venus and the Earth itself. □

H. J. Melosh is in the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

1. Eugster, O. *Science* **245**, 1197–1202 (1989).
2. McSweeney, H. Y. *Rev. Geophys.* **23**, 391–416 (1985).
3. Gratz, A. J., Nellis, W. J. & Hinsey, N. A. *Nature* **363**, 522–524 (1993).
4. Vickery, A. M. thesis, State Univ. New York, Stony Brook (1984).
5. Melosh, H. J. *Icarus* **59**, 234–260 (1984).
6. Melosh, H. J. *Geology* **13**, 144–148 (1985).
7. Binzel, R. P. & Xu, S. *Science* **260**, 186–191 (1993).
8. McKinnon, W. B. *Nature* **363**, 211–212 (1993).
9. Hofmann, F. *Ecolae Geol. Helv.* **66**, 83–100 (1973).
10. Hofmann, B. & Hofmann, F. *Ecolae Geol. Helv.* **85**, 788–789 (1992).

NUCLEIC ACID ARCHITECTURE

Tetrads through interdigitation

Dinshaw J. Patel

It is 40 years since the classical anti-parallel duplex structure of DNA was revealed¹, but there seems to be no end to the surprising topologies that nucleic acids can adopt. The latest example appears on page 561 of this issue², where Gehring, Leroy and Guéron describe the solution structure of a new four-stranded structural motif, designated the 'i-motif'. The motif results from the antiparallel interdigitation of protonated cytosine-cytosine ($\text{C}\cdot\text{C}^+$) parallel-stranded duplexes, an arrangement that is unexpected and thought-provoking.

Both high-resolution nuclear magnetic resonance (NMR) and single-crystal X-ray techniques have contributed to the recent burst of activity in investigation of the higher-order structure of nucleic acids. The result has been determination of the solution structures of pyrimidine-purine-pyrimidine³ and purine-purine-pyrimidine⁴ triplexes, as well as crystallographic⁵ and solution^{6–10} structures of guanine (G) quadruplexes with a range of strand directions. These structures have characteristics — base-stacking patterns, hydrogen-bonded pairing alignments, strand directions, groove dimensions and cation-coordination sites — that go beyond our previous knowledge of right-handed¹¹ and left-handed¹² anti-parallel DNA duplexes.

The formation of $\text{C}\cdot\text{C}^+$ base pairs in a parallel-stranded duplex was initially defined from X-ray diffraction analysis of polycytidylic acid fibres¹³, and later confirmed from single-crystal X-ray stud-

ies of the deoxy($\text{C}\cdot\text{G}$)-d($\text{C}\cdot\text{G}$) duplex¹⁴ and NMR studies of DNA oligomer duplexes containing the d($\text{C}\cdot\text{G}\cdot\text{A}$)-d($\text{C}\cdot\text{G}\cdot\text{A}$) segment¹⁵, all undertaken at acidic pH. What was unexpected was the mutual intercalation of parallel-stranded $\text{C}\cdot\text{C}^+$ duplexes to form an i-tetrad motif at acidic pH in solution.

Gehring *et al.*² undertook an NMR study of d(T-C-C-C-C) and related deoxycytidine-rich sequences at acidic pH, and readily detected unusual distance-dependent nuclear Overhauser enhancement (NOE) patterns indicative of a novel structural motif. Equilibrium dilution experiments established that d(TC_5) forms a tetramer at acidic pH. The four strands were shown to be equivalent in the NMR spectrum, and the $\text{C}\cdot\text{C}^+$ pairing alignment was monitored by following the exchangeable downfield shifted hydrogen-bonded deoxycytidine imino proton (one per $\text{C}\cdot\text{C}^+$ pair), as well as the resolved hydrogen-bonded and exposed deoxycytidine amino protons.

These data were consistent with formation of parallel-stranded duplexes aligned through $\text{C}\cdot\text{C}^+$ pairing. The authors detected unusually strong NOEs between sugar H1' protons on non-adjacent deoxycytidines in the d(TC_5) sequence. This observation serves as a fingerprint for the i-tetrad motif because sugar-sugar proton NOEs are not detected in duplexes, triplexes or quadruplexes because in these structures there is more than 5 Å between protons on adjacent sugars. The striking conclusion from all this was that pairs of

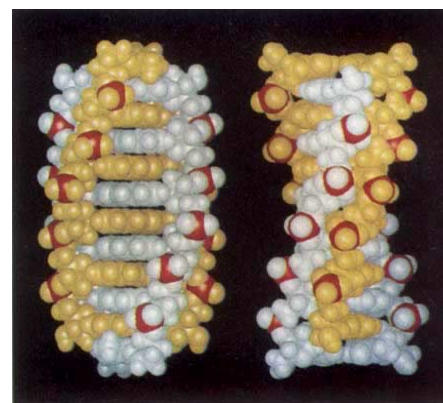


FIG. 1 Two views of the i-tetrad structure of d(TC_5). One of the parallel-stranded duplexes in the tetraplex is shown in yellow and the other in white; the backbone phosphates are in red. Left, a view of the wide groove, showing the mutual intercalation of the $\text{C}\cdot\text{C}^+$ base pairs from the two parallel-stranded duplexes in the tetraplex. Right, a view of the narrow groove where adjacent phosphates from anti-parallel oriented strands are positioned opposite each other.

parallel-stranded $\text{C}\cdot\text{C}^+$ duplexes were interdigitating in an anti-parallel orientation, such that deoxycytidines far apart in the sequence were brought close together in the i-tetrad structure.

The solution structure of the d(TC_5) tetraplex, determined following molecular dynamics calculations, exhibits the following features (Fig. 1). Individual $\text{C}\cdot\text{C}^+$ parallel-stranded duplexes are right-handed and underwound (twist angle about 16°) with intercalation sites generated at every base pair along the helix (Fig. 2). Two parallel-stranded duplexes are aligned in an anti-parallel orientation to form the tetraplex with alternate stacking of the $\text{C}\cdot\text{C}^+$ base pairs from the two duplexes. The helix axis exhibits dyad symmetry, with successive

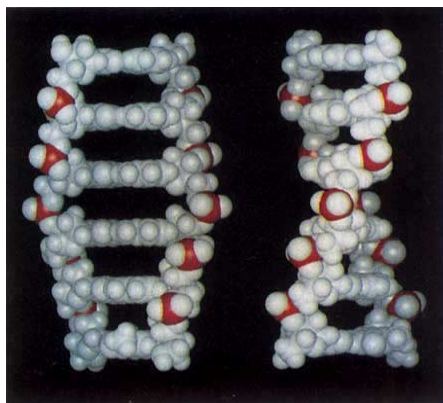


FIG. 2 Two views of one of the two parallel-stranded $\text{C}\cdot\text{C}^+$ duplexes in the i-tetrad structure of d(TC_5). Note that the intercalation sites are generated at every base pair and that the helix is unwound (twist angle about 16°) but retains its right-handed helical twist. The two views differ by a rotation of 90° along the helix axis.

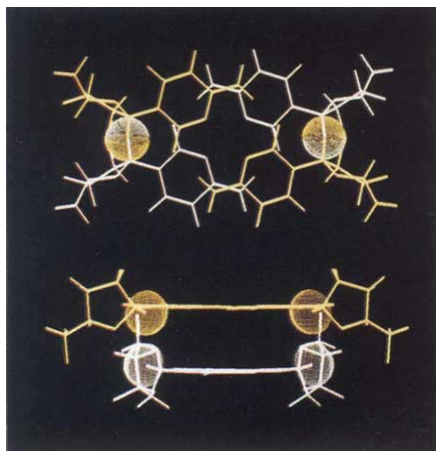


FIG. 3 Top, A view down the helix axis showing the stacking of adjacent C-C⁺ base pairs originating from the two parallel-stranded duplexes (yellow and white) in the i-tetrad of d(TC₅). The sugar H1' protons (dotted balls) on adjacent C-C⁺ base pairs are directly over each other and would be expected to, and do, show strong NOEs to each other. Bottom, A view normal to the helix axis, once again showing the close proximity of sugar H1' protons on adjacent C-C⁺ base pairs.

base pairs related by a helical rise and a rotation. The tetraplex structure has a pair of wide grooves (shortest phosphate-phosphate separation between duplexes of 15.5 Å; Fig. 1, left) and a pair of narrow grooves (shortest phosphate-phosphate separation between duplexes of 4.9 Å; Fig. 1, right) with the backbone phosphates being in close proximity across the groove in the latter case.

Other features are extensive pair-wise sugar-phosphate backbone interactions with close van der Waals sugar-sugar contacts between the anti-parallel aligned strands from each of the C-C⁺ parallel-stranded duplexes across the minor groove in the tetraplex structure. In addition, adjacent C-C⁺ base pairs whose long axes are orthogonal to each other, are stacked face-to-face (Fig. 3, top). The deoxycytidine N³ nitrogens are maximally separated, while the carbonyl and amide dipoles exhibit a reverse orientation between successive base pairs. The H1' protons of adjacent base pairs are directly over each other (Fig. 3, top and bottom) and emphasize that the observed pattern of NOEs reflects the order of base-pair stacking rather than the covalent sequence.

In a companion paper¹⁶ the same group describes monitoring of the exchange characteristics of the imino and amino protons of the hemiprotonated C-C⁺ base pairs in the d(TC₅) tetraplex. Added catalyst has no effect on the imino proton exchange rates, reflecting the intrinsic exchange catalysis contributions¹⁷ by the acceptor atom in the C-C⁺ base pair. These studies show that the base-pair opening kinetics for the i-tetrad are slow,

with lifetimes of minutes. Such lifetimes are comparable to those reported for a few base pairs in the core of transfer RNA, but are distinctly longer than the few milliseconds observed for Watson-Crick base pairs in anti-parallel duplexes¹⁸. These long lifetimes must reflect the face-to-face and fully intercalated alignment of the hemiprotonated C-C⁺ base pairs. Further, the absence of exchange broadening at the deoxycytidine amino protons accounts for the symmetry of the C-C⁺ base pair because it means that the imino proton must switch between the N³ atoms of the deoxycytidines within the pair at a rate of over 80,000 per second.

The characterization of an i-tetrad structure provokes conjecture as to the role of this structural motif in biology. What, for instance, is the effect of supercoiling in inducing i-tetrad formation in repeating riboC_n and deoxyC_n sequences? Can antibodies to the i-tetrad motif be generated, and can proteins that recognize and bind to it be isolated? Further, is it possible to generate i-tetrads at neutral pH through interactions with either nucleic acids or proteins? These and other questions remain to be investigated, and one eagerly awaits structure-function correlations involving this intriguing variation on nucleic-acid architecture. Lastly, speculation on the biological significance of the i-tetrad motif is prompted by the repetitive deoxycytidine stretches, such as deoxy (C₄A₂), which are found in telomeres. The deoxycytidines of four consecutive stretches could fold together into an i-tetrad motif, with the deoxyguanines of the partner strand forming a G-tetrad. □

Dinshaw J. Patel is in the Program in Cellular Biochemistry and Biophysics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.

1. Watson, J. D. & Crick, F. H. C. *Nature* **171**, 737-738 (1953).
2. Gehring, K., Leroy, J. L. & Guéron, M. *Nature* **363**, 561-565 (1993).
3. Radhakrishnan, I., Patel, D. J., Veal, J. M. & Gao, X. *J. Am. chem. Soc.* **114**, 6913-6915 (1992).
4. Radhakrishnan, I. & Patel, D. J. *J. Am. chem. Soc.* **115**, 1615-1617 (1993).
5. Kang, C. H., Zhang, X., Ratliff, R., Mayzis, R. & Rich, A. *Nature* **356**, 126-131 (1992).
6. Wang, Y. *et al.* *J. molec. Biol.* **222**, 819-832 (1991).
7. Smith, F. W. & Feigon, J. *Nature* **356**, 164-168 (1992).
8. Wang, Y. & Patel, D. J. *Biochemistry* **31**, 8112-8119 (1992).
9. Cheong, C. & Moore, P. B. *Biochemistry* **31**, 8406-8414 (1992).
10. Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S. & Bolton, P. H. *Biochemistry* **32**, 1899-1904 (1993).
11. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. *Nature* **286**, 567-573 (1980).
12. Wang, A. H. J. *et al.* *Nature* **282**, 680-686 (1979).
13. Langridge, R. & Rich, A. *Nature* **198**, 725-728 (1963).
14. Cruse, W. B. *et al.* *Biochemistry* **22**, 1833-1839 (1983).
15. Robinson, H., van der Marel, G. A., van Boom, J. H. & Wang, A. H. J. *Biochemistry* **31**, 10510-10517 (1992).
16. Leroy, J. L., Gehring, K., Kettani, A. & Guéron, M. *Biochemistry* **23**, 6019-6031 (1993).
17. Guéron, M., Kochoyan, M. & Leroy, J. L. *Nature* **328**, 89-92 (1987).
18. Guéron, M. & Leroy, J. L. in *Nucleic Acids and Molecular Biology* Vol. 6 (eds Eckstein, F. & Lilley, D. M.) 1-22 (Springer, New York, 1992).

DAEDALUS

Cellular fabric

ONE of the neatest of natural polymerizations is the condensation of glucose to cellulose. The reverse hydrolysis also goes very easily. Strong acid will do it, though cellulose-digesting bacteria and fungi use enzymes like cellulase. Daedalus points out that enzymes, however subtle, are simply catalysts. They must facilitate the forward and back reactions equally well. So he reckons you could make cellulose simply by adding cellulase to a solution of glucose. As fast as cellulose forms it would precipitate from the solution, driving the equilibrium to make more of it. Growing plants seem to thicken their cell walls in just this way.

Cellulose by itself is fairly feeble stuff. It shows its real strength only when incorporated into some composite material. Wood, for example, is a composite of cellulose and lignin. Rot fungi digest the cellulose, leaving a shrunken lignin skeleton. So, says Daedalus, by immobilizing cellulase on rotten wood and treating it with glucose solution, you could reverse the process. The wood would accumulate internally distributed cellulose until it had regained its original size and strength. Brief heating could then denature the cellulase and stabilize the product.

This ingenious process should even work *in situ*. Decaying roofs or rotting floorboards could be restored with a dash of enzyme and a jug of syrup! On a smaller scale, the technique might also restore worn and threadbare garments. But Daedalus goes further. He recalls the wonderful strength and tenacity of nature's organic/inorganic composite materials. Bone, ivory, mother-of-pearl, all have an inorganic matrix filled with organic polymer. His process could induce this subtle microstructure in any solid porous structure.

So he is now filling a wide range of porous solids with cellulose. Chalk, concrete or plaster should be transformed into dense, tough, durable new engineering materials. Cellulose-metal composites would hold oil splendidly, and would damp out mechanical noise. Light and porous woods, such as balsa and spruce, could themselves be reinforced, upgrading them to the density and strength of teak or lignum vitae. Even cellulose-reinforced cellulose should be possible. A cotton fabric stretched taut and filled with cellulose should give a sort of prestressed leather.

Daedalus's cellulose composites will transform engineering. They will even be welcomed by the green lobby. For, of course, they will be biodegradable. At the end of their long road, the fungi will lie in wait

David Jones

Sound and rock art

SIR — Neither the location nor subject matter of prehistoric rock art has yet been satisfactorily explained¹. Quantitative data from sites sampled in France, together with experience on two other continents, provide a possible explanation for the location and subject matter of Upper Palaeolithic parietal art.

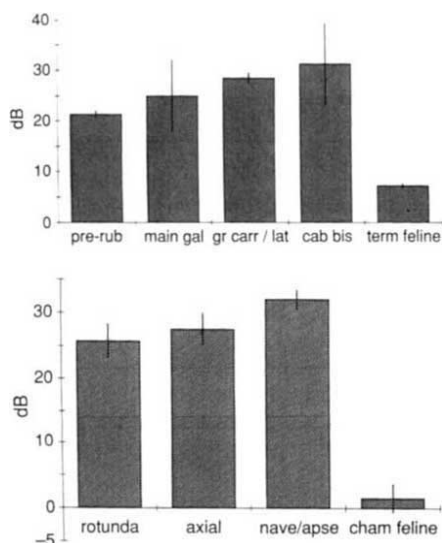
Open-air art sites as well as deep decorated caves from the Upper Palaeolithic give reflected sound levels significantly above ambient. For example, at the open-air site of l'Oreille d'Enfer, echoes were measured at 53 ± 3 dB against a background of 45 ± 2 dB ($P < 0.0003$). Further, in the deep caves of Font-de-Gaume and Lascaux, the images of horses, bulls, bison and deer are found in regions with high levels of sound reflection, whereas feline art is found in regions of the caves with poor acoustics (see figure). The difference between reflected sound levels at ungulate versus feline art locations is statistically significant at $P < 0.0004$ within Font-de-Gaume and $P < 0.0001$ within

Lascaux. These results suggest an acoustic influence on both the placement and content of the art.

It has been previously observed that the shape of a cave exerted some general influence on the placement of species². Indeed, shape is one major determinant of cave acoustics. However, the highly sound-reflecting axial gallery decorated with ungulates and the acoustically dead chamber of felines in the same cave of Lascaux are both narrow dead-end tunnels, suggesting that the cave shape was influential only to the extent that it does affect the acoustics.

Echoes have been mentioned as a phenomenal attribute of certain rock art sites³, and a correspondence has been suggested between deep cave painting placement and locations that resonate at particular musical notes⁴. But no objective quantitative acoustical data have been reported, nor is there a comprehensive theory that understandably relates acoustics to the rock art content.

Statistical tabulations of the content of European Palaeolithic parietal art² can be reinterpreted as showing that more than 90% of the subjects depicted fall into the biological classes Artiodactyla, Perissodactyla and Proboscidea,



Comparison of reflected sound levels in different regions of Palaeolithic caves. Experiments with a sound-burst device were recorded using a PMD420 Marantz cassette recorder with a Dynamic LOZ Shure SM57 omnidirectional microphone, and analysed using a Brüel and Kjær model 2232 precision sound-level meter. The intensity of reflected sound is expressed on the value axis as decibels above background noise levels. Error bars represent one standard deviation of two to four replicate tests. Abbreviations are: pre-rub, pre-rubicon (entrance tunnel); main gal, main galleries; gr carr/lat, grand carrefour (intersection) and lateral gallery; cab bis, cabinet de bisons; term feline, terminal fissure with feline; rotunda (hall of bulls); axial, axial gallery; nave/apse, nave/apse area; and cham feline, chamber of felines. Top, Font-de-Gaume; bottom, Lascaux.

collectively grouped as ungulates. Why might prehistoric humans have chosen to make images of hooved mammals in sound-reflecting environments?

It is known that some ancient cultures considered echoing a supernatural phenomenon. Experimentation with the sound reflection at rock art sites reveals that percussion noises (from clapping or producing stone tools) can yield echoes that sound similar to the galloping of a horse, and that reverberation of percussion noises can sound like the thundering of a buffalo stampede. This phenomenon of hoofbeat-like echoes thus relates to the ungulate images as well as to the sound-reflecting canyons and caves where the art is typically located.

These measurements and observations suggest that Palaeolithic ungulate art was produced in response to percussive sound reflections perceived as hoofbeats. The production of hoofbeats via sound reflection could have been part of a ritual intended to summon up game.

Steven J. Waller

109 Saint Regis Drive,
Newark, Delaware 19711, USA

1. Bahn, P. G. & Vertut, J. *Images of the Ice Age* (Facts on File, New York, 1988).
2. Leroi-Gourhan, A. *Treasures of Prehistoric Art* (Abrams, New York, 1967).
3. Steinbring, J. *Am. Indian Rock Art* **17**, 102–113 (1992).
4. Reznikoff, I. & Dauvois, M. *Bull. Société Préhist. Français* **85**, 238–246 (1988).

Relations of fish and tetrapods

SIR — The evolutionary transition from fishes to tetrapods, which occurred about 400 million years ago, has come under intense scrutiny in the light of new molecular data and reinterpretation of the morphological evidence^{1–4}. There is little disagreement that tetrapods are most closely related to the lobe-finned fishes (sarcopterygians), whose members possess fin supports resembling tetrapod limb bones, and have, in at least some taxa, functional lungs. But there is debate over the relationship of the tetrapods to the two extant sarcopterygian lineages, the lungfishes and the coelacanth.

Although the traditional view^{5,6} supports the coelacanth as closest relative, some recent morphological analyses^{7,8} have favoured the lungfishes, or lungfishes + coelacanth, as sister group of the tetrapods. Previous molecular data^{1,3,9} have supported either the coelacanth or the lungfishes, and this question is generally considered to be unresolved². We have sequenced approximately three kilobases of mitochondrial DNA, including the entire 12S ribosomal (r) RNA and 16S rRNA genes, from the coelacanth and all three lineages of lungfishes. Phylogenetic analyses including all available vertebrate sequences for

these genes support the lungfishes as the closest living relatives of the tetrapods.

We sequenced the genes for 12S rRNA, tRNA^{Val}, and 16S rRNA in *Latimeria chalumnae* (coelacanth, tissue courtesy of J. A. Musick), *Lepidosiren paradoxa* (South American lungfish), *Protopterus* sp. (African lungfish) and *Neoceratodus forsteri* (Australian lungfish, tissue courtesy of W. Bemis). We also sequenced a portion of the mitochondrial cytochrome *b* gene in *L. paradoxa* and *N. forsteri* for comparison with published data for that region in the other sarcopterygian fishes. The sequenced regions correspond to sites 648–3,229 (RNA genes) and 14,842–15,124 (cytochrome *b*) in the published human sequence. For phylogenetic analyses, we included all complete vertebrate sequences for these genes available in databases. Carp, a ray-finned fish (*Cyprinus carpio*), was used as the outgroup. The available tetrapods are frog (*Xenopus laevis*), chicken (*Gallus gallus*), human (*Homo sapiens*), seal (*Phoca vitulina*), cow (*Bos taurus*), whale (*Balaenoptera physalus*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*).

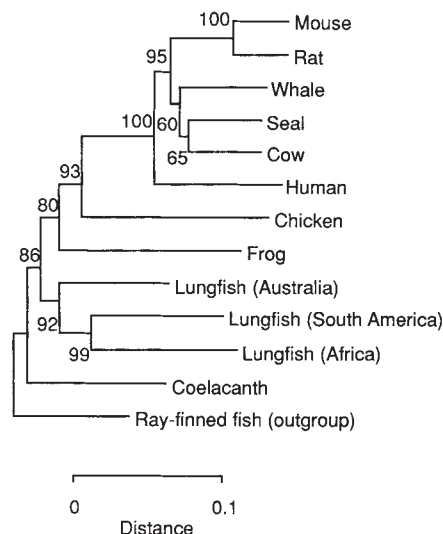
Two different methods of phylogenetic analysis, neighbour-joining and maximum parsimony, resulted in trees that

Phylogenetic tree of tetrapods and lobe-finned fishes based on mitochondrial DNA sequences of an approximately 3-kb region of the mitochondrial genome, including the entire genes for 12S rRNA, tRNA^{Val} and 16S rRNA, and a 283-bp portion of the cytochrome *b* gene. The neighbour-joining method was used with the Kimura distance to correct for different rates of substitution in transitions versus transversions, and for multiple hits. A total of 2,073 sites, of which 830 are variable, could be aligned (sites with deletions not used). Bootstrap P-values (10), from 0–100%, are indicated on the tree. A maximum parsimony analysis (not shown) resulted in a single most-parsimonious tree which also supported a lungfish–tetrapod relationship. DNA was extracted from liver and muscle samples stored at -70°C , amplified by PCR, and both complementary strands sequenced^{14,15} with the use of 23-oligonucleotide primers. For cytochrome *b*, we used two pairs of primers, each spanning the same region in *Lepidosiren* and *Neoceratodus*. The programs NJBOOT2 and Treeview (K. Tamura, Pennsylvania State University) were used to construct the neighbour-joining trees and perform the bootstrap analysis (2,000 replications). Our cytochrome *b* sequence for *Lepidosiren* differs slightly from that obtained elsewhere⁹ but both sequences, used separately, resulted in identical topologies. The programs Metree and Treeshow (A. Rzhetsky, Pennsylvania State University) were used to test the significance (from

support a lungfish–tetrapod relationship (see figure). Three statistical tests^{10,11} of this node on the phylogenetic tree gave similar results: bootstrap P-value (86%), branch-length significance (82%) and rejection of alternative topologies (89%). The same neighbour-joining topology was obtained using either a simple p-distance (proportion), a Jukes–Cantor distance to correct for multiple hits, or a Kimura distance to correct for transition bias and multiple hits.

The three lungfishes from different continents form a monophyletic group, with the South American and African species as sister taxa. Three already well-established groups, Tetrapoda, Amniota and Mammalia, are each monophyletic. This is the largest molecular dataset to address the question of the closest living relative of tetrapods, and these results support the findings of previous mitochondrial DNA studies that included fewer sites and taxa^{3,9}. A lungfish–tetrapod relationship is further supported by enzyme usage in the synthesis of urea¹². The phylogeny derived from the biochemical and molecular evidence provides an independent framework for interpreting the morphological changes involved in the transition from fishes to tetrapods.

Identification of the specific lineage of living fishes that is our closest relative is also important for understanding the evolution of non-fossilizable tetrapod traits and the genetics of tetrapod limb



zero) of internal branches in the neighbour-joining tree, and to compare the total sum of branch lengths of that tree with the length of alternative topologies¹¹. The new sequences reported here have been deposited in EMBL (Z21921-3, 21926-8). (A copy of the complete alignment has been deposited as supplementary information in the London editorial office of *Nature*; that, and other technical information, is available from S. B. H.)

development¹³. Although most molecular evidence strongly favours a lungfish–tetrapod relationship, DNA sequence data from additional slow-evolving genes should provide an increasingly robust phylogenetic framework for understanding vertebrate evolution.

S. Blair Hedges

Carla Ann Hass

Linda R. Maxson

Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

- Gorr, T., Kleinschmidt, T. & Fricke, H. *Nature* **351**, 394–397 (1991).
- Forey, P. L. *Nature* **351**, 347–348 (1991).
- Meyer, A. & Dolven, S. I. *J. molec. Evol.* **35**, 102–113 (1992).
- Marshall, C. R. & Schultze, H. P. *J. molec. Evol.* **35**, 93–101 (1992).
- Romer, A. S. *Vertebrate Paleontology* (Univ. Chicago Press, 1966).
- Schultze, H. P. in *Origins of the Higher Groups of Tetrapods* (eds Schultze, H. P. & Trueb, L.) 29–67 (Comstock, Ithaca, New York, 1991).
- Chang, M.-M. in *Origins of the Higher Groups of Tetrapods* (eds Schultze, H. P. & Trueb, L.) 3–28 (Comstock, Ithaca, New York, 1991).
- Forey, P. L., Gardiner, B. G. & Patterson, C. in *Origins of the Higher Groups of Tetrapods* (eds Schultze, H. P. & Trueb, L.) 145–172 (Comstock, Ithaca, New York, 1991).
- Meyer, A. & Wilson, A. C. *J. molec. Evol.* **31**, 359–364 (1990).
- Felsenstein, J. *Evolution* **39**, 783–791 (1985).
- Rzhetsky, A. & Nei, M. *Molec. biol. Evol.* **9**, 945–967 (1992).
- Mommsen, T. P. & Walsh, P. J. *Science* **243**, 72–75 (1989).
- Tabin, C. & Laufer, E. *Nature* **361**, 692–693 (1993).
- Hedges, S. B., Bezy, R. L. & Maxson, L. R. *Molec. biol. Evol.* **8**, 767–780 (1991).
- Hedges, S. B., Bogart, J. P. & Maxson, L. R. *Nature* **356**, 708–710 (1992).

Weak queen or social contract?

SIR — There may be a simpler explanation for the experiment that Reeve and Nonacs reported as evidence for social contracts in wasps¹. Reeve and Nonacs observed increased aggression towards dominant alpha queens by beta queens when eggs were experimentally removed from the nest, and they interpret this as evidence for retaliation against alpha queens for eating eggs laid by the beta queen. Another interpretation consistent with these results is that the beta queen infers from the empty brood cells that the alpha queen is weak and might be vulnerable to a challenge. This would result in the observed escalated aggression by the beta queen and interference with the alpha queen's attempts to lay eggs, particularly by beta queens who are nearly as big as the alpha queen. This hypothesis also fits with West-Eberhard's finding² that alpha queens limit the reproduction of beta queens by restricting access to empty cells, as only the alpha queen begins new cells.

The weak-queen hypothesis is simpler than the social contract hypothesis because the former assumes only that the beta queen perceives that the cells are empty, and therefore that there is a problem with the alpha queen. The social contract hypothesis assumes that the beta queen perceives that the cells are empty, recognizes that some of those cells contained her eggs, and that she further assumes that the alpha female ate the eggs. Furthermore, she fails to recognize that most of the eggs the alpha queen ate she also laid. The weak-queen hypothesis also accords better with what we know about egg recognition. Eggs are thought to be recognized for only the brief time during which they are guarded^{2,3}. In this species eaten eggs were on average 11.4 minutes old, and females guarded their eggs for up to 20 minutes².

Another advantage of the weak-queen hypothesis is that it provides a more natural explanation of why the beta and not the alpha queen increases aggression. The alpha queen would not increase aggression as she is already the dominant queen. Under the social contract hypothesis, the alpha queen encountering empty cells formerly containing eggs could also think that the social contract had been broken by the beta queen. It would then follow that the alpha queen should also increase aggression when eggs are experimentally removed.

The beta queen was observed by Reeve and Nonacs to increase aggression towards the alpha queen in July and not

in June. The social contract hypothesis explains this difference as resulting from the timing of laying of reproductive eggs, which is more likely to be in July than in June. According to this interpretation, beta queens are selected to care about the parentage of eggs destined to become reproductives and not of eggs destined to become workers. One difficulty with this interpretation is that it relies on the untested assumption that workers can predict the eventual caste of eggs in the face of unpredictable future effects of individual nest conditions^{1,4,5}. There is another, simpler interpretation. Early in the season females fight to establish rank so absolute levels of aggression are high, and differences in aggression among females are small. Therefore there is not much room for experimental change in aggression levels. Later in the season, as rank becomes better established, the alpha and beta queens diverge more in behaviour, ovarian development and hormone titres^{6,7}. This provides more room for an increase in aggression by beta queens in an experiment. The beta queen's behaviour changes less in June because in June hormone titres of the beta female are still high and more similar to those of the alpha queen^{2,8}. In July, when provided with the stimulus to lay eggs (empty cells²), more of the hormone that governs both egg-laying and aggression (juvenile hormone) is produced, resulting in increased aggression.

Removal of pupae did not elicit aggression by the beta queen. However, pupae are often naturally removed in attempts to rid the brood of parasites such as *Chalcoela iphitalis*, whose larvae can move from cell to cell, destroying all pupae⁹. This makes it unlikely that empty pupal cells would be an indication of a weak queen, so this result is consistent with both hypotheses.

In sum, although the social contract hypothesis is intriguing and possibly true, the weak-queen hypothesis explains all the observed phenomena in a simpler way.

Joan E. Strassmann

Department of Ecology and
Evolutionary Biology,
Rice University, PO Box 1892,
Houston, Texas 77251, USA

REEVE AND NONACS REPLY — We disagree that the weak-queen hypothesis is more parsimonious than our social contract hypothesis¹; more important, several results strongly contradict the former.

First, the weak-queen hypothesis assumes that the beta queen perceives that the cells are empty, assumes that empty cells reflect an unspecified 'weakness' in the alpha queen, with beta failing to remember that these cells contained eggs immediately before both

queens were removed from the nest and continuing to perceive this 'weakness' despite normal behaviour by the alpha queen. Thus both hypotheses are complex. Strassmann argues that our hypothesis cannot easily explain the observation that queens usually stop guarding their own eggs and refrain from eating their nestmate's eggs only minutes after such eggs are laid². However, a restricted period of guarding or egg eating does not imply that queens cannot discriminate between their own and their nestmate's eggs beyond this period. Indeed, nest-switching experiments suggest that queens can discriminate among eggs of varying relatedness to them hours to days after such eggs have been laid¹⁰.

Second, if the weak-queen hypothesis were correct, we should have seen at least some increase in aggression by the beta when worker-destined eggs were removed (as this also creates empty cells and thus should indicate a weak alpha). In fact, beta's aggression dropped slightly after removal of such eggs. Strassmann argues that the lack of increase in beta's aggression resulted only from beta's high juvenile hormone titres early in the summer, which makes betas nearly as aggressive as alphas, by contrast to later in the summer when betas are less aggressive and thus capable of displaying greater net increases in aggression. But this explanation cannot apply to our results since natural rates of beta aggression in our early (worker-egg) colonies (10.4 acts per hour) were not different from those of later (reproductive-egg) ones (10.2), probably because dominance hierarchies had stabilized even in the former. These data preserve the evolutionary hypothesis that selection favours enhanced protectiveness of eggs that are likely to be reproductive-destined late in the season.

According to the weak-queen hypothesis, beta's aggression should be relatively low when alpha tries to lay eggs in treatment nests, because this should indicate that alpha is less afflicted by the unspecified 'weakness'. In fact, beta is more aggressive when alpha tries to lay eggs.

We have devised a critical test of the hypotheses. According to the weak-queen hypothesis, an alpha that replaces eggs at a high rate after egg removal treatments should be viewed by beta as less afflicted by the weakness than an alpha that replaces eggs at a low rate, and thus the former alphas should be treated less aggressively than the latter. According to the social-contract hypothesis, betas should be more aggressive to queens that lay replacement eggs at higher rates, as it is these alphas that will be perceived as most flagrantly 'cheating' on the social contract. The result is decisively in favour of the social con-

tract. In treatment colonies, betas exhibited larger increases in aggression toward alphas that laid replacement eggs at high rates (+16.9 acts per hour) than toward alphas that laid replacement eggs at low rates (-7.4; $P < 0.05$).

The weak-queen hypothesis is interesting in its own right and should be tested in other contexts, but it can be rejected as an explanation for our results. Social wasp queens indeed seem to maintain their social contract over reproduction by close monitoring of relative reproduction coupled with the threat of aggressive retaliation against cheating.

Hudson Kern Reeve

Museum of Comparative Zoology,
Harvard University,
26 Oxford St,
Cambridge,
Massachusetts 02138, USA

Peter Nonacs

Department of Zoology,
University of California,
Davis, California 95616, USA

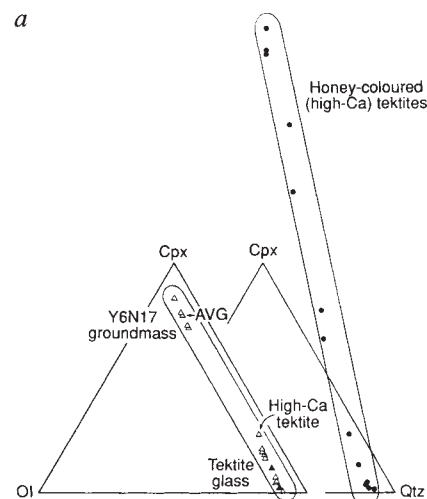
1. Reeve, H. K. & Nonacs, P. *Nature* **359**, 823-825 (1992).
2. West-Eberhard, M. J. *Misc. Publ. Mus. Zool. Univ. Mich.* **140**, 1-101 (1969).
3. Noonan, K. M. in *Natural Selection and Social Behavior* (eds Alexander, R. D. & Tinkle, D. W.) 18-44 (Chiron, New York, 1981).
4. Solis, C. R. & Strassmann, J. E. *Funct. Ecol.* **4**, 531-541 (1990).
5. Strassmann, J. E. *J. Kansas ent. Soc.* **62**, 353-362 (1989).
6. Röseler, P.-F., Röseler, I. & Strambi, A. *Insectes Sociaux* **27**, 97-107 (1980).
7. Turillazzi, S., Marino Piccioli, M., Hervatin, L. & Pardi, L. *Monitore zool. ital. (N.S.)* **16**, 75-88 (1982).
8. Strassmann, J. E. *Florida Ent.* **64**, 74-88 (1981).
9. Strassmann, J. E. *Ecology* **62**, 1225-1233 (1981).
10. Klahn, J. E. & Gamboa, G. J. *Science* **221**, 482-484 (1983).

K/T melt glasses

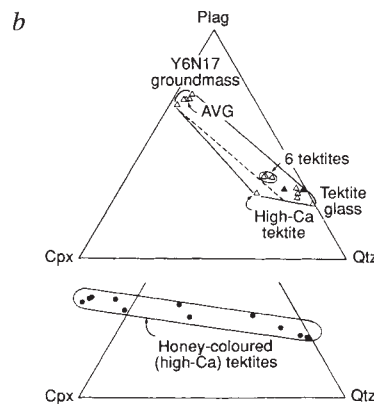
SIR — In our recent paper¹, we used the chemical composition of a melt rock from the Chicxulub structure to confirm Chicxulub's identity as an impact crater and as the source of impact melt spherules found at the Cretaceous/Tertiary (K/T) boundary in Haiti. We did this by first demonstrating that the quenched molten portion of the melt rock does not lie on igneous fractionation paths and thus is not likely to be a volcanic product. Second, we showed that this composition lies on a chemical mixing trend seen among K/T boundary impact melt glasses, and thus concluded that both sets of samples were probably produced by the same impact event. Although these results have since been substantiated by other groups using different methods²⁻⁴, Schuraytz and Sharpton⁵ in Scientific Correspondence suggested that our results were fortuitous for reasons we address here.

Schuraytz and Sharpton⁵ drew attention to the potential effects of hydrothermal alteration on the composition of the groundmass. We considered these:

a, Projected compositions of honey-coloured (high-Ca) tektites⁸, in addition to those of the Chicxulub melt rock and Haitian tektites originally used in ref. 1, projected onto the cpx–ol–qtz plane. The honey-coloured (high-Ca) tektites appear to define a different chemical trend (right) from that seen among more siliceous (generally grey to black) varieties of tektites (left). The intersection of these trends is responsible for the curvature of the data plotted in Fig. 1 of ref. 5, where the



compositions of both groups of tektites were combined. b, The same compositions shown in a, but projected onto the cpx–plag–qtz plane. The original data set used in ref. 1 (top) is consistent with the mixing trend inferred from the cpx–ol–qtz projection and again we note that the subset of high-Ca tektites appears to define a different chemical trend (bottom) from that seen among more siliceous varieties of tektites. The dispersion of the Chicxulub groundmass data is due solely to the heterogeneous distribution of groundmass phases in a single thin section and should not be used to infer large-scale mixing trends among impact-melted material.



indeed, our laboratory was the first to describe the extent of these effects on the mineralogy and texture of the melt rock and overlying polymict breccia^{1,6,7}; in the case of the melt rock, they are manifested as quartz and possibly anhydrite veins. Schuraytz and Sharpton suggest that sodic feldspar in the groundmass is an additional non-isochemical alteration product. However, the groundmass is a pristine-looking intergranular assemblage without any reaction textures that would support their supposition. Sodic feldspar does not appear to be forming at the expense of calcic plagioclase, nor is there any reaction between these minerals and the more potassic feldspar that occurs in coronas surrounding clasts in the rock. The petrographic characteristics of the groundmass do not change with distance from the veins crosscutting the rock, which are the obvious avenues for chemical exchange. Furthermore, none of the other phases (chlorite and/or smectite, epidote, quartz, pumellyite, prehnite and other zeolites) usually associated with the type of alteration processes envisioned by Schuraytz

and Sharpton is present in the groundmass. Thus, we feel it is reasonable to interpret the groundmass as either a primary (but possibly metastable) crystallization assemblage or an isochemical devitrification product in which one or both feldspars crystallized from residual glass.

Schuraytz and Sharpton also point out the utility of projecting the data onto the cpx–ol–qtz and cpx–plag–qtz diagrams; we had done this before submitting our paper for publication, and found the same mixing trend to be evident (see figure). As Schuraytz and Sharpton suggest, the heterogeneous distribution of groundmass phases within a single thin section can lead to scatter in the inferred groundmass composition; for this reason, we use the average groundmass composition to define the mixing trends in the figure. The primary difference between our results and those of Schuraytz and Sharpton is the additional data regarding very Ca-rich tektite compositions. We agree that this subset of Haitian tektites diverges from the mixing trend that we observed among more siliceous varieties (see figure), but suggest that these different groups of tektites represent ejecta that flowed through and was excavated from different regions of the crater.

Because the formation of the Chicxulub melt sheet involved several different lithologies which were mechanically disaggregated, melted and mixed over distances of several to tens of

kilometres, there is no *a priori* reason to suspect that their melt rock sample (C1N10) and ours (Y6N17) would have the same composition, particularly when they were separated laterally by ~35 km and also appear to represent different melt horizons. Sample Y6N17 is a relatively clast-rich sample with a microcrystalline groundmass which most likely represents the rapidly cooled upper margin of a melt sheet, whereas the clast-free and medium-grained C1N10 sample⁴ appears to represent a deeper and more homogenized horizon. Because Y6N17 cooled so quickly, it is unlikely to have been affected by post-impact differentiation processes, although this may have occurred elsewhere in the melt sheet. Although we agree with Schuraytz and Sharpton that albitization can affect groundmass compositions, this cannot be the explanation of the differences in the compositions of the samples: the groundmass of Y6N17 contains more CaO than does C1N10, not less, and even complete $\text{NaSi} \rightleftharpoons \text{CaAl}$ or $\text{Na} \rightleftharpoons \text{K}$ exchange cannot produce the groundmass composition of Y6N17 from C1N10, nor the trend seen between the siliceous tektites and Y6N17. Consequently, our initial observation, that the composition of the groundmass represents a melt which lies off the igneous fractionation paths associated with volcanic suites of rock (Fig. 1a of ref. 1), is still true and implies that the melt was produced by impact processes.

Although we feel that post-impact hydrothermal alteration processes have not significantly affected the texture and major-element composition of Y6N17, they have clearly affected the texture and mineralogy of the overlying polymict breccia, as we have described previously^{6,7}, and it appears that a later event (or events) disturbed the argon isotopic system^{3,4}. We hope that future studies will lead to better understanding of these alteration processes and the extent to which they affected the evolution of the crater, but it already appears that these processes were localized and did not affect all samples uniformly.

David A. Kring

William V. Boynton

Lunar and Planetary Laboratory,
University of Arizona,
Tucson,
Arizona 85721, USA

1. Kring, D. A. & Boynton, W. V. *Nature* **358**, 141–144 (1992).
2. Hildebrand, A. R. et al. *Lunar Planet. Sci.* **XXIII**, 539–540 (1992).
3. Swisher, C. C. III et al. *Science* **257**, 954–958 (1992).
4. Sharpton, V. L. et al. *Nature* **359**, 819–821 (1992).
5. Schuraytz, B. C. & Sharpton, V. L. *Nature* **362**, 503–504 (1993).
6. Kring, D. A., Hildebrand, A. R. & Boynton, W. V. *Lunar Planet. Sci.* **XXII**, 755–756 (1991).
7. Hildebrand, A. R. et al. *Geology* **19**, 867–871 (1991).
8. Izett, G. A. *J. Geophys. Res.* **96**, 20879–20905 (1991).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

Sex, genes and stereotypes

Steven Rose

The Sexual Brain. By Simon LeVay. MIT Press: 1993. Pp. 168. \$22.50, £14.95. UK publication date, 1 July.

SIMON LeVay is a neuroanatomist whose early research, with David Hubel and Torsten Wiesel at Harvard Medical School, on neural plasticity in the visual system attracted considerable notice within the neuroscience community. But in 1991, by then at the Salk Institute, he sprang to much wider fame with a paper in *Science* in which, on the basis of the morphometric analysis of brains from a sample of 19 declared homosexual male AIDS victims, 16 presumed heterosexual men and 6 presumed heterosexual women, he reported that the size of a nucleus in the medial preoptic region of the hypothalamus varied with sexual orientation among the men.

"Gay brains" immediately joined "gay genes" as the focus for intense debate both in the straight popular press (the *New York Times*, *Time* and *Newsweek* included) and in the gay community itself*. It is far from irrelevant in this context that LeVay himself is a declared gay, for the issues raised by the question of the possible biological origins of differences in sexual orientation seem to differ from those posed by other examples of biological determinism, of genes 'for' violence, criminality or differences in IQ scores between ethnic groups. The difference lies not so much in the nature of the logical chain supposed to link a particular biological measure with a particular behaviour or character trait, for this presents similar complexities in each case, but in the responses to the deterministic claim. Significant sections of the gay community have actively welcomed the LeVay report, regarding it as demonstrating that gays are indeed biologically different and should therefore be neither morally stigmatized nor socially penalized. If gays are born, not made, as one gay activist has put it to me, then parents can relax; gay school-teachers are not going to be able to corrupt the morals of otherwise heterosexual children. Other gays have been vehemently critical of any such suggestion of biology as destiny.

The Sexual Brain begins with a brief trip through current thinking on the biology, evolution and development of sex, with an overview of the neural mechanics of copulation, before focusing on its main agenda, the biological and, above all, the

brain determinants of sexual activity and orientation. Unsurprisingly, the book culminates in a discussion of LeVay's own findings and recent twin studies and their implications for understanding the nature and origins of homosexuality (mainly male; lesbianism is mentioned briefly though it doesn't even merit an index

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Homosexuality — born or bred?

entry) and transsexuality. The book's pleasingly informal style marks it as being intended for a wide audience, which its theme and author's reputation will undoubtedly ensure, although anyone unfamiliar with brain structures will find it hard going in places, not least because, astonishingly for a book written by an anatomist, it relies entirely on verbal description, with not a single diagram or photograph to help the uninitiated through neuroscience's dog-Latin.

LeVay's own data and the related genetic claims have been subject to intense critical scrutiny over the past couple of years (in, for instance, Anne Fausto-Sterling's *Myths of Gender* (Basic Books, 1992) and Ruth Hubbard and Elijah Wald's *Exploding the Gene Myth* (Beacon, 1993)). Without citing his critics directly, LeVay carefully acknowledges some of the interpretative problems his observations present. But he is much less cautious — and this is a really worrying feature of a book intended for a wide audience — in accepting the data and claims of others that seem to support the thrust of his argument. His goal is clear: to rebut what he sees as a "Freudian" or psychobiological account of the origins of homosexuality which root it in early childhood experience and to replace it with a

biological alternative. Fancy, for LeVay (who has a penchant for Shakespearean epigrams to head his chapters), is clearly bred in the head — if not the genes. To this end he cheerfully extrapolates from rats to humans and sweeps together a *mélange* of distinctly dubious anecdotal and anthropological data and sociobiological speculations, from stories of young twins playing naked in the woods so that they might attract homosexually inclined truck drivers to molest them, to E. O. Wilson's bizarre fantasy that 'gay genes' could have adaptive value (if, despite presumably diminishing the number of offspring their possessors might leave, they also predispose towards assisting in the rearing of siblings' offspring). To LeVay, it seems, everything in this mix is of equal evidential weight.

But the sharpest problem, to my mind, is the way in which his approach reifies homosexuality and needlessly polarizes 'social' and 'biological' causation. He frequently argues as if 'homosexuality' and 'heterosexuality' were fixed properties of an individual rather than descriptors of parts of a continuum of developmentally expressed polyvalent sexual behaviour and activity that vary during any individual's lifetime and have ascribed social and cultural meanings that differ widely from age to age and society to society. I, for one, find it difficult to believe that being homosexual in San Francisco today has the same meaning for an individual, either biologically or socially, as it might have done in Victorian England. To be sure, LeVay nods in this interactionist direction, as in his closing words: "Like waterlilies, we swing to and fro with the currents of life, yet our roots moor us each to our own spot on the river's floor." This is a distinctly more agreeable image than Wilson's of genes holding culture on a long leash, but the dominant motif remains that of fixity, rootedness, over which the social currents gently wash. Yet in an earlier incarnation LeVay would not, I suspect, have chosen this image to delimit the richness of his own findings in the much less complex field of the interplay of neural specificity and plasticity, of ontogenetic experience, in shaping the cellular responses of the developing visual cortex. Understanding the origins and significance of our own sexuality and sexual preferences will require at least as careful an integration of theory, experiment and observation as does interpreting the visual system of the cat. □

Steven Rose is in the Brain and Behaviour Research Group, Open University, Milton Keynes MK7 6AA, UK. His most recent book, *The Making of Memory* (Bantam, 1992), has won this year's Rhône-Poulenc science book prize (see last week's issue).

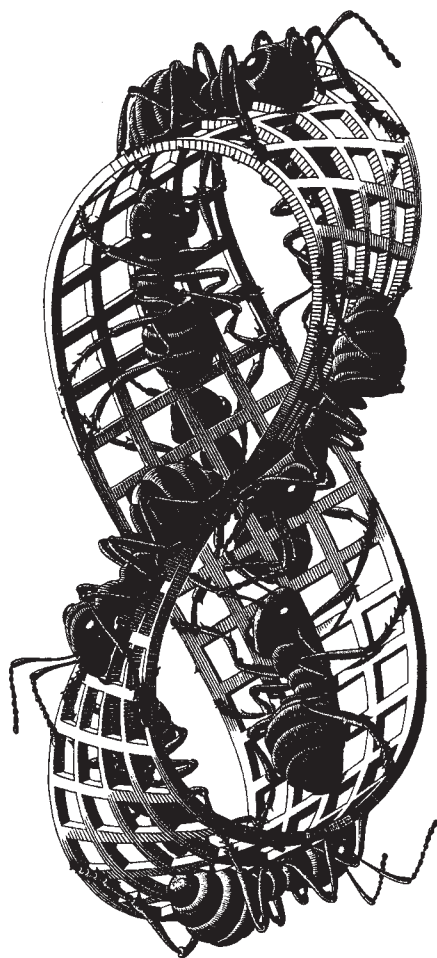
* See also the leading News and Views article, "Is homosexuality hard-wired?" (*Nature* 353, 13, 1991).

Strip-tease

Desmond King-Hele

Möbius and his Band: Mathematics and Astronomy in Nineteenth-Century Germany. Edited by J. Fauvel, R. Flood and R. Wilson. Oxford University Press: 1993. Pp. 172. £19.50, \$29.95.

THE real subject of this book is best exposed by slyly altering the title to 'Möbius and his band of fellow-German



From Möbius and his Band

Möbius Strip II by M. C. Escher (1963).

mathematicians and astronomers'. We are presented with six essays on German mathematics and astronomy in the nineteenth century, as the subtitle correctly indicates. The essays are loosely centred on Möbius, but "this is not a biography of Möbius", as the first sentence of the preface declares. The catchy title *Möbius and his Band* is not only misleading but also wrong, because the essay on topology by Norman Biggs shows that the Möbius band is not really his, having been described a little earlier by J. B. Listing, a pioneer of topology and pupil of Gauss.

All this may seem carping criticism, but it is symptomatic of a sense of confusion and repetition that permeates the book. The preface boldly makes a virtue of

necessity, admitting that Möbius was "not outstanding in any particular way, but was a serious, competent, professional scholar" and was therefore "a good mirror of his time". The trouble is that the six contributors have to refer to Möbius from time to time, in deference to the plan of the book, and they often find only a blank. For example, Ian Stewart, in his wide-ranging survey of "Möbius's modern legacy" remarks that "historical recognition is at best a fickle thing, but in this case its absence is appropriate. . . . To put it bluntly, he was a bit of a plodder." (In the United Kingdom today he would probably be one of the many unemployed scholars.) Allan Chapman, at the end of his essay on the development of astronomy in Germany and elsewhere during Möbius's lifetime, outlines the blank more politely (or perhaps sarcastically?), by saying that Möbius "would have been intimately acquainted with the developments discussed above. . . in which his own observatory would have played a part". Certainly it would have been bizarre if he had been unaware of progress in astronomy; and Möbius was not a bizarre man.

If Möbius seems to be 'the man who wasn't there', where was he? Born in 1790 near Leipzig, August Möbius became a student at Leipzig University in 1809, and in 1816 was appointed 'extraordinary professor' in astronomy. This was a "lowly form of academic life", as John Fauvel remarks in his introductory essay: it meant that he could charge fees for lecture courses. As "he was not an especially charismatic teacher", students apparently only "came to his courses when he advertised them as free". After 28 years, in 1844, he became a full professor, and in 1848 director of the Leipzig Observatory. He described the Möbius band in 1858 and died in 1868. Although he made little impact in astronomy, apart from writing two popular books, he was a good mathematician and wrote three notable books.

The first of these, in 1827, was on the 'barycentric calculus', the main subject of Jeremy Gray's engaging essay on "Möbius's geometrical mechanics". Möbius helped in pioneering several fruitful concepts in projective geometry, including duality and homogeneous coordinates. His name is also attached to a function, an inversion formula, a transformation and a net, all clearly explained by Gray. His other two books were on celestial mechanics (1843) and the theory of circular transformations (1855). These were substantial achievements, and helped to transform him into a man who was there: certainly he was there at Leipzig University between 1809 and 1868.

Even if some intellectual defects lurk within, the book looks good: it sports an attractive multicoloured jacket, and is well illustrated, with many photographs of places and instruments, and a nice gallery

of engravings of German mathematicians. There are also numerous mathematical diagrams, all clear and relevant, and the many mathematical equations and expressions are well set out. It is the text that has deficiencies, in particular too many repetitions. The meagre known details of Möbius's life are inevitably repeated *ad nauseam*, but the repetition extends to more peripheral matters: on page 3 we are told that Napoleon's defeat of Prussia at Jena in 1806 sparked a Prussian cultural renaissance; the same story is repeated on page 10, and again on page 23. These repetitions would have been picked up if there had been a proper index, but in fact the 'index' is only a list of personal names; it is therefore difficult to guess which essay will cover a particular topic.

When all this is said, however, the book is a useful history of some aspects of German astronomy and mathematics in the nineteenth century, especially as seen in relation to modern advances. □

Desmond King-Hele was formerly deputy chief scientific officer at the Royal Aircraft Establishment, Farnborough, UK.

Air raid

Len Goodwin

Disease Transmission by Insects: Its Discovery and 90 Years of Effort to Prevent It. By James R. Busvine. Springer: 1993. Pp. 361. DM98, \$69, £38.50.

AIR travel adds a new dimension of mobility to the earthbound, enhancing speed, range and precision; it is easy to see that when insects evolved, parasitic organisms benefited from being carried on the feet and in the guts of flying arthropods to new environments and hosts. It is more difficult to imagine how, as a result of small random mutations, some of them could have developed their highly sophisticated life cycles, using a particular species of insect as an intermediate host.

Most parasites live at peace with their natural hosts, doing them little harm. But when a new, susceptible animal invades their territory they may kill, and human intrusion into the wilderness has led to epidemics of yellow fever, trypanosomiasis, malaria and many other lethal diseases. It took a long time before the idea that diseases might be carried by insects was taken seriously, but over a score of years around 1900, causative organisms and their vectors were identified and their life cycles elucidated.

James Busvine recounts the stories of these discoveries and of the pioneers — of many nations — who made them, and pays tribute to the scientists' ingenuity, imagination and courage. They were fully

aware of the dangers; and through accident or deliberate exposure in testing a theory, many suffered and some died. Among the less well-known stories is that of typhus fever, the terrible disease of besieging armies, beleaguered cities, filthy camps and prisons, carried by body lice. The name of the infective organism, *Rickettsia prowazeki*, was conferred by Rocha-Lima, a Brazilian who contracted typhus and became very ill; both of the scientists who are celebrated in the name, an American and an Austrian, died.

The insect vector appeared at first to be an easy target for the control of disease, and there were some notable early successes such as the campaigns against yellow fever and malaria in Cuba and Panama. But the detection and elimination of breeding sites was expensive and never-ending, and the available agricultural pesticides were toxic or too irritant for human use. The discovery of new, potent, synthetic compounds was full of promise.

Busvine, at the London School of Hygiene and Tropical Medicine, has been working with the application of these compounds for most of his life. As long ago as 1940 he was testing the efficacy of new substances as delousing agents by spraying them on the underpants of London vagrants, and soon afterwards was impregnating shirts with the new Swiss insecticide DDT for British soldiers serving overseas (a measure that saved thousands of lives). He describes the worldwide impact of DDT and deals systematically with methods for controlling mosquitoes, midges, flies, ticks, mites, bugs, lice and fleas. DDT and other chlorinated hydrocarbons eventually became ineffective through misuse and the development of resistance, and new substances with different modes of action (anticholinesterases and synthetic pyrethroids) superseded them. These are more expensive than DDT and insects develop resistance to them also. The old methods — environmental control, traps, repellents and nets — are back in fashion, and some new biological and genetic control techniques are on trial. Future success depends on international training, cooperation and finance.

Busvine's style is easy and conversational, and he presents a wealth of specialized knowledge and personal experience to the general reader. With the excuse that this is not a book of reference, there is a list of contents but no index — which is a pity because there is plenty worth referring to. The illustrations are rather grey and flat, and there are some irritating misprints — perhaps because printing was done in faraway Bangalore. □

Len Goodwin, formerly with the Wellcome Trust, is at Shepperlands Farm, Park Lane, Finchampstead, Wokingham, Berkshire RG11 4QF, UK.

Seeds of Gaia

Bill Chaloner

Faith in a Seed: The Dispersion of Seeds and Other Late Natural History Writings by Henry Thoreau. Edited by B. P. Dean. Island: 1993. Pp. 283. \$25.

WHEN Henry David Thoreau's aunt asked him on his deathbed whether he had made peace with his god, he replied that he did not know that they had quarrelled. It is disappointing to find that a man who can come up with such impromptu wit *in extremis* can write such uninspired ramblings when he is in a more relaxed frame of mind.

Thoreau died at the age of 44, apparently from tuberculosis aggravated by a cold caught while walking in the rain in his



beloved Massachusetts woods. He is probably best known for *Walden*, written while living in a hut on the edge of a lake outside Boston in the years 1845–47, in which he expounded his rugged New England philosophy and agri-ecological contemplation. Thoreau is noteworthy for his meticulous observation of plant and animal life in last-century New England, and seeing in its workings a moral message for humankind. After the appearance of *Walden* in 1854 he published little more, but continued to make voluminous notes on the biology of the woodlands near his home in Concord. These were intended to lead to a book on the dispersion of seeds, which he left unfinished. Research into these notes became the subject of a PhD dissertation at the University of Connecticut, by the editor of this posthumous work, Bradley P. Dean. His edited version of Thoreau's notes, and those for three other small unpublished natural-history works of Thoreau, form the basis of this book. Beyond Thoreau's text, there is the delightful addition of Abigail Rorer's beguiling little marginal drawings, as shown here. These are not only a delight in themselves, but give a pleasingly nineteenth-century touch to the page layout, as the lines are tailored to wander around a red-oak seedling or the skeleton of a deer-mouse.

In his foreword, Gary Nabhan tells us

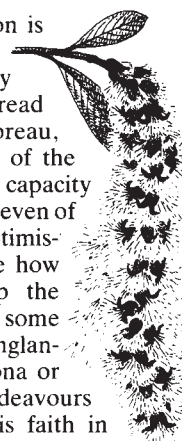
with disconcerting honesty that Odell Shepherd, editor of the *Heart of Thoreau's Journals*, considered Thoreau's observations on seed biology "a chapter of his lifework best left unpublished". Readers must judge for themselves, but I think it is at least fair to say that those who look here for new scientific insights into the reproductive biology of plants will be disappointed. The work is of value more for what it tells us about Thoreau than about the plant life of Massachusetts.

Thoreau's somewhat anecdotal style is reminiscent of Gilbert White's *The Natural History of Selborne*, rather than that of his contemporary Charles Darwin. The comparison is of interest, since Thoreau read and wrote notes on *The Origin of Species* in 1860, while writing his seed book. He was at that time also acting as a collector for Louis Agassiz, king of Harvard biology and arch-opponent of Darwinism. Indeed, Thoreau was invited to sit on a Harvard committee charged with an annual evaluation of its Natural History syllabus — *plus ça change!* Since Thoreau was evidently an exponent of evolution by natural selection, the minutes of those meetings would make interesting reading.

Thoreau seems to have been an early victim of the conflict between the Two Cultures. He wrote in 1851: "what sort of science is that which enriches the understanding but robs the imagination?" He need not have worried; Gary Nabhan believes that "the loveliest passages in his treatise on seeds are those in which his deep reservoir of literary knowledge wells up into his science". That is a fair comment, and one that would surely have pleased Thoreau.

The title of this compilation is not merely apt but descriptive of its contents; like many expositions of faith, it can be read at different levels. For Thoreau, seeds are a symbol not only of the dissemination of life and its capacity for regeneration, but perhaps even of resurrection. He suggests, optimistically, that "it is remarkable how Nature endeavours to keep the earth clothed with wood of some kind". (There writes a New Englander; the inhabitants of Arizona or Alaska might see Nature's endeavours in other terms . . .) But his faith in Nature goes further: "the very earth itself is a granary and a seminary . . . so that to some minds its surface is regarded as the cuticle of one living creature". In the ubiquity and fecundity of seeds, Thoreau seems to have found intimations of Gaia a good century before Lovelock. □

Bill Chaloner is in the Department of Biology, Royal Holloway, University of London, Egham Hill, Egham, Surrey TW20 0EX, UK.



Dangerous diagnostics

Frank Barnaby

World Inventory of Plutonium and Highly Enriched Uranium 1992. By David Albright, Frans Berkhout and William Walker. Oxford University Press: 1993. Pp. 246. £25, \$39.95.

THIS volume provides the only comprehensive assessment of how much plutonium and highly enriched uranium — the basic ingredients of nuclear weapons — exists in the world today, and where and in which form they are to be found. Given the crucial importance of this information for global security, the *World Inventory* is essential reading for all interested in international affairs and an indispensable reference book for researchers.

The authors analyse in detail the inventories in the nuclear-weapon states (the United States, the former Soviet Union, the United Kingdom, France and China), the *de facto* nuclear-weapon states (Israel, India and Pakistan), countries suspected of having ambitions to produce nuclear weapons (Iraq, North Korea, Iran and Algeria) and countries "backing away from nuclear weapons" (Argentina, Brazil, South Africa and Taiwan). The surprising omission of South Korea from the last group is not explained.

Some of the most important data in the *World Inventory* is given in the accompanying box. The authors estimate that the error margins on their figures, by far the best nongovernmental ones available, are $\pm 15\%$ for plutonium (Pu) and $\pm 30\%$ for highly enriched uranium (HEU). Most uncertain is the size of military inventories in the former Soviet Union. This is a serious matter. Some 25 tonnes of ex-Soviet military Pu cannot be accounted for. There is particular concern that when ex-Soviet nuclear weapons are dismantled in Russia and the Pu in them transferred from military to civil control, some of the Pu will be stolen and sold on the black market. Weapon-grade Pu is worth at least a million dollars a kilogram on the black market. Given the very poor salaries of Russian scientists and technicians, the temptation to sell Pu must be almost irresistible.

Whereas the world's stocks of military Pu and HEU are roughly constant, amounts of civil Pu are increasing steadily. According to present plans, about 310 tonnes of civil Pu will be separated by the year 2000. Of this, about 60 tonnes will be stored in the United Kingdom, about 50 tonnes in Japan, about 40 tonnes in each of Germany and Russia, and about 15 tonnes in France. Unilateral and bilateral disarmament agreements may lead to the

dismantlement during the 1990s of 30,000 or so US and ex-Soviet nuclear weapons containing about 90 tonnes of Pu and about 450 tonnes of HEU.

The *World Inventory* stimulates one to consider how Pu and HEU should be disposed of. Discarding HEU is fairly straightforward. By mixing HEU with enough natural or depleted U to reduce the concentration of ^{235}U from 90 to 2 or 3%, it can be used as fuel in civilian light-water power reactors. The United States has agreed to buy 500 tonnes of HEU from Russia for this purpose.

The disposal of Pu is not so easy. Its use as reactor fuel — for breeder or ordinary reactors — is not economic. The electricity produced by commercial breeder reactors will be relatively expensive until the price of uranium increases considerably. This will not happen for decades. Mixed oxide fuel for light-water reactors is expensive — about twice the cost of ordinary uranium oxide fuel.

Other suggested methods for getting rid of Pu include firing it into the Sun using rockets, transmuting it into other elements in special reactors or particle accelerators, 'burning' it in fast (non-breeding) reactors, and mixing it with

high-level radioactive waste. The risk that a rocket might accidentally fall back to Earth with its Pu payload is environmentally unacceptable. Machines for transmuting large amounts of Pu and suitable fast reactors have not yet been developed. The best way of dealing with Pu would be to incorporate it in silica glass with high-level radioactive waste and to place it in a geological depository.

Until the world's Pu is permanently disposed of, it should be owned, stored and safeguarded by the International Atomic Energy Agency. The international management of Pu would be much better than the current chaotic situation of national ownership and storage and give some confidence that Pu was not being diverted to military uses. The safe storage of HEU is perhaps even more important. As Luis Alvarez has reminded us, the background neutron rate in weapon-grade HEU is so low that terrorists would have a good chance of setting off a high-yield nuclear explosion by simply dropping 30 kg of the material onto another 30 kg. □

Frank Barnaby is at Brandreth, Station Road, Chilbolton, Stockbridge, Hampshire SO20 6AW, UK.

World amounts of Pu and HEU

■ **Total amounts.** The amounts of plutonium (Pu) and highly enriched uranium (HEU) in the world's civil and military inventories are huge — currently about 1,100 tonnes of Pu and 1,400 tonnes of HEU (excluding roughly 150 tonnes of HEU in naval reactors and their fuel cycles). About a quarter of Pu and almost all the HEU is used for military purposes, mostly in nuclear weapons. HEU consists mainly of weapon-grade uranium containing at least 90% of the fissile isotope ^{235}U . The world's nuclear arsenals contain about 50,000 nuclear warheads, about 95% of them US and ex-Soviet. On average, each warhead contains about 3 kg of Pu (weapon-grade Pu contains at least 93% ^{239}Pu) and 15 kg of HEU.

■ **Civil plutonium.** The inventory of civil Pu is increasing at about 62 tonnes a year, contained in the roughly 9,300 tonnes of spent fuel discharged annually by the world's nuclear power reactors. Of the world's 780 or so tonnes of civil Pu, about 81% is still in spent reactor fuel; about 135 tonnes have been separated from spent reactor fuel. Of the separated Pu, about 86 tonnes is kept in store; the rest is in the breeder reactor fuel cycle and in mixed oxide fuel for ordinary (thermal) reactors (in mixed oxide fuel, plutonium oxide is mixed with uranium oxide). Of the Pu in store, about 35% is in Russia and about 47% is in the United Kingdom.

■ **Military plutonium.** Of the world's 260 or so tonnes of military Pu, about 70% is in warheads; the rest is stored outside warheads. Most is in the United States (about

112 tonnes) and the former Soviet Union (about 125 tonnes). The United Kingdom is estimated to have about 11 tonnes of military Pu, France about 6 tonnes, and China about 2.5 tonnes. The best estimate for the amount of military Pu produced so far by Israel is about 250–450 kg. The figure for India is about 330 kg. The Israeli figure is questionable. According to the information given by Mordechai Vanunu, the Israeli whistle-blower, the Israelis probably produced at least 500 kg up to 1986. They may have produced about 300 kg since. There is no reason to doubt Vanunu's figures, which are based on first-hand knowledge, and it is hard to understand why the authors have done so.

■ **Highly enriched uranium.** The bulk (97%) of the world's HEU is held by the United States and the former Soviet Union. There are about 40 tonnes of HEU in the United Kingdom, France and China put together. Pakistan is estimated to have produced between 200 and 300 kg of HEU. The amount of HEU produced by South Africa is very uncertain, with a potential inventory of between 200 and 525 kg. This presents a serious challenge to the International Atomic Energy Agency to verify independently South Africa's declaration of its HEU inventory. According to President F. W. de Klerk, South Africa produced only six nuclear weapons. The authors' estimate indicates that South Africa had enough HEU to produce between 10 and 25 nuclear weapons. The cynic will wonder whether a few have been hidden.

In situ measurements constraining the role of sulphate aerosols in mid-latitude ozone depletion

D. W. Fahey*, S. R. Kawa*†, E. L. Woodbridge*†, P. Tin*†, J. C. Wilson‡, H. H. Jonsson‡, J. E. Dye§, D. Baumgardner§, S. Borrmann§, D. W. Toohey||, L. M. Avallone¶||, M. H. Proffitt*†, J. Margitan#, M. Loewenstein**, J. R. Podolske**, R. J. Salawitch††, S. C. Wofsy††, M. K. W. Ko‡‡, D. E. Anderson§§, M. R. Schoeberl|||, K. R. Chan**

* Aeronomy Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

† Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, Colorado 80309, USA

‡ University of Denver, Department of Engineering, Denver, Colorado 80208, USA

§ National Center for Atmospheric Research, Boulder, Colorado 80307, USA

|| University of California, Irvine, Department of Geosciences, Irvine, California 92717, USA

¶ Harvard University, Department of Chemistry, Cambridge, Massachusetts 02138, USA

Jet Propulsion Laboratory, Pasadena, California 91109, USA

** NASA Ames Research Center, Moffett Field, California 94035, USA

†† Harvard University, Department of Earth and Planetary Sciences, Cambridge, Massachusetts 02138, USA

‡‡ Atmospheric and Environmental Research, Inc., Cambridge, Massachusetts 02139, USA

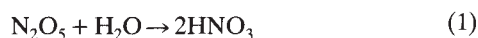
§§ Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland 20723, USA

||| NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

***In situ* measurements of stratospheric sulphate aerosol, reactive nitrogen and chlorine concentrations at middle latitudes confirm the importance of aerosol surface reactions that convert active nitrogen to a less active, reservoir form. This makes mid-latitude stratospheric ozone less vulnerable to active nitrogen and more vulnerable to chlorine species. The effect of aerosol reactions on active nitrogen depends on gas phase reaction rates, so that increases in aerosol concentration following volcanic eruptions will have only a limited effect on ozone depletion at these latitudes.**

In polar regions in winter, ozone depletion occurs primarily through the catalytic action of chlorine liberated by reactions on polar stratospheric cloud (PSC) particles^{1,2}. At low to mid-latitudes, greater solar illumination and the absence of PSCs cause the destruction rate of ozone to be determined by a combination of catalytic cycles³⁻⁵. Gas-phase models of the atmosphere show that the dominant destruction cycle in these regions involves nitrogen oxides, not halogen species. Laboratory studies have shown, however, that the nitrogen oxides N_2O_5 and $ClONO_2$ react on the surface of sulphuric acid solutions with compositions similar to those of stratospheric aerosol particles⁶⁻⁸.

aerosol



Sulphate aerosol particles, formed from biological and volcanic sources of sulphur, are present throughout the lower stratosphere⁹. For stratospheric water vapour and temperature conditions, the reaction probability of (1) is nearly independent of temperature⁸. In contrast, the reaction probability of (2) increases exponentially with decreasing temperature, becoming equal to that of reaction (1) below 200 K (ref. 8). Both reactions decrease NO_x ($= NO + NO_2$) and increase HNO_3 within the reactive nitrogen reservoir, NO_y (see Fig. 1). Reducing NO_x changes the balance among the various photochemical cycles that produce and destroy ozone³⁻⁵. For example, production of the inactive chlorine reservoir species $ClONO_2$ and HCl is reduced, thereby increasing ClO and its associated ozone loss cycles¹⁰⁻¹².

Observations of NO_y species, available primarily from remote

surroundings in the stratosphere, have provided important evidence for the effect of reaction (1) in background conditions at both mid- and high latitudes^{6,9,13-18}. At mid-latitudes, measurements from the ATMOS experiments^{5,17} are the most comprehensive to date in the lower and middle stratosphere. Photochemical models for the species in Fig. 1 show better agreement with ATMOS data when reaction (1) is included, but significant discrepancies remain in the NO comparison in the lower stratosphere where measurement uncertainties are greatest. To clarify the role of reactions (1) and (2), we present here recent *in situ* measurements of NO and NO_y along with aerosol surface area (SA), O_3 and ClO . The measurements are made simultaneously at mid-latitudes in the lower stratosphere. With these measurements, the way in which the photochemical balance within the NO_y reservoir depends on aerosol SA can be evaluated more precisely than with current remote sounding data. In the evaluation, the NO_x/NO_y ratio is formulated for both background and enhanced aerosol conditions and compared to representative model calculations. For background conditions, the observed ratio is found to be lower than that predicted using only gas-phase reactions whereas it agrees favourably when reaction (1) is included. This agreement provides important support for the role of reaction (1) in the lower stratosphere. Inclusion of reaction (2) in the model comparison results in no further change in the NO_x/NO_y ratio under these conditions, consistent with laboratory results for mid-latitude temperatures. With the N_2O_5 sulphate reaction, the importance of nitrogen oxides in ozone depletion at mid-latitudes is considerably reduced in models of the background atmosphere whereas that of Cl_y and HO_x is increased³⁻⁵. Furthermore, the N_2O_5 sulphate reaction lessens the impact on ozone of additional NO_x from stratospheric aircraft in these models, but increases the sensitivity of ozone to

growth in halogen species^{3,19}.

For enhanced aerosol SA following major volcanic eruptions^{20–22}, further reductions in the NO_x/NO_y ratio are expected as the rates of reactions (1) and (2) increase^{4,23,24}. Unchecked reductions in NO_x within the NO_y reservoir would lead to catastrophic ozone loss rates as ClO is liberated from the inactive ClONO_2 reservoir⁴. These loss rates could be similar to those that occur in the Antarctic stratosphere in spring as a result of PSC reactions and denitrification. The formation rate of N_2O_5 is expected to limit the increase in the rate of reaction (1) as aerosol SA increases. Because N_2O_5 is formed predominantly by the reaction of NO_2 and NO_3 at night (see Fig. 1), the formation rate falls quadratically with decreases in available NO_x . Thus, increases in aerosol SA lead to further decreases in N_2O_5 , but changes in NO_x become negligible. This saturation of the NO_x/NO_y reduction is expected from reaction (1) only as aerosol SA increases well above background values⁴. Saturation of reaction (2) is expected only at much higher aerosol SA because ClONO_2 is formed whenever ClO and NO_2 are present. Thus, reaction (2) at high aerosol is the primary cause of catastrophic ozone loss rates found by models⁴.

The eruption of Mount Pinatubo in June 1991 provided an excellent opportunity to evaluate enhanced aerosol conditions^{20,22}. The eruption injected three times more sulphur mass into the stratosphere than the eruption of El Chichón in 1982.²⁴ Presented here are *in situ* measurements made before and after the arrival of sulphate aerosol to northern mid-latitudes in 1991–92 that show a nonlinear decrease in NO_x/NO_y with increasing surface area, suggesting saturation. The change agrees well with model predictions incorporating reaction (1) for aerosol values up to 40 times background. In this case, the addition of reaction (2) to the model reduces the predicted NO_x/NO_y by only a small factor at the highest aerosol values. The agreement further validates the role of reaction (1) in the stratosphere and demonstrates that photochemical changes resulting from large observed increases in sulphate aerosol are limited at mid-latitudes, consequently limiting expected changes in ozone loss rates.

Data collection and model calculations

The data presented here were acquired as part of the NASA Airborne Arctic Stratospheric Expedition II (AASE II) during the period September 1991 to March 1992. *In situ* measurements were made with the NASA ER-2 aircraft up to 20 km altitude for a wide variety of species. These measurements are used to derive the ratios NO_x/NO_y (see Fig. 2) and ClO/Cl_y as indices of the partitioning of the reactive nitrogen, NO_y , and reactive chlorine, Cl_y , reservoirs. Here we focus on data acquired close to local noon at mid-latitudes on 17 September 1991 and 22 March 1992. September flight data in Fig. 2 are unique in AASE II because background aerosol conditions ($\text{SA} < 1 \mu\text{m}^2 \text{cm}^{-3}$)⁹ are found at low solar zenith angles at mid-latitudes. Figure 3 displays data from the March and September flights selected to contrast air masses that have comparable photochemical conditions but large differences in aerosol abundance.

FIG. 1 Diagram denoting the principal species of the NO_y reservoir ($= \text{NO} + \text{NO}_2 + 2\text{N}_2\text{O}_5 + \text{HNO}_3 + \text{ClONO}_2 + \text{HO}_2\text{NO}_2 + \dots$) and reaction pathways. The NO_x/NO_y ratio within the reservoir is set by the balance of many reactions involving O_x , HO_x , Cl_y and Br_x species. The thickness of the arrows is nominally proportional to the conversion rate between component species⁸. The effect of the N_2O_5 and ClONO_2 reactions with H_2O on sulphate aerosol is to reduce the steady-state abundance of NO_x within NO_y . N_2O_5 is produced mainly at night in the recombination of NO_2 and NO_3 , and destroyed during the day by photodissociation, reforming NO_2 . NO_3 is formed mainly at night in the reaction of NO_2 with O_3 . ClONO_2 is formed in the recombination of ClO and NO_2 . HNO_3 , the reaction product of (1) and (2) and principal NO_y species, re-forms NO_x during the day through photolysis. In addition, increases in reactive chlorine in the form of ClO and in the hydroxyl radical, OH, are expected for aerosol SA at or above background values^{10,11,19}.

TABLE 1 NO_x/NO_y values

	Observations* (Background aerosol)	Gas phase†	LSS model* Heterogeneous‡
$[\text{OH}] \times 1$	0.12	0.18	0.10
$[\text{OH}] \times 2$	—	0.14	0.09
$[\text{OH}] \div 2$	—	0.24	0.11
JPL J_{HNO_3} §	—	0.19	—

* Average values for the measurement interval in Fig. 2 from 42° to 48° latitude corresponding to background aerosol SA.

† $[\text{OH}]$ values used are 0.00062 p.p.b.v. at noon with a diurnal average of 0.00017 p.p.b.v.

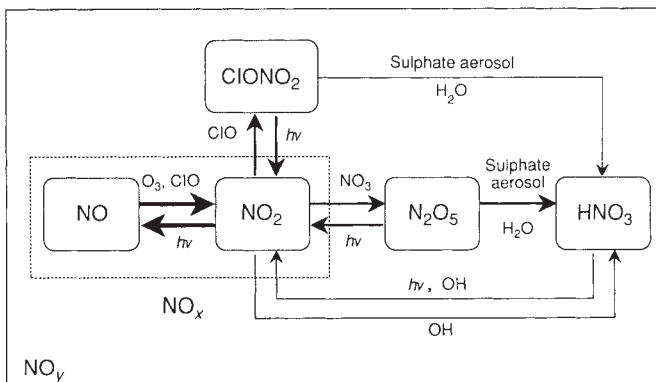
‡ $[\text{OH}]$ values used are 0.00076 p.p.b.v. at noon with a diurnal average of 0.00022 p.p.b.v.

§ NO_x/NO_y value calculated using the photolysis rate of HNO_3 , J_{HNO_3} , derived from JPL-recommended cross-sections at 298 K (refs 8, 26). All other calculations use J_{HNO_3} derived from more recent cross-section measurements^{50,51}.

A local steady-state (LSS) model is used here to calculate NO_x/NO_y along the flight track (see Fig. 2). The LSS calculation requires the specification of seven parameters: date, latitude, temperature, pressure, O_3 , ClO, OH and photolysis rates. Here all parameters are known or measured in the aircraft, except the latter two. OH is interpolated along the flight track from a separate full diurnal model calculation that is parameterized in a general way using the aircraft data²⁵. Photolysis rates are calculated using radiation-scattering models²⁶ and the results of laboratory studies⁸. The LSS model was compared to several full diurnal photochemical models using identical input parameters and was found to agree to better than 10% at 20 km altitude for both the gas-phase and heterogeneous cases using reactions (1) and (2). As the LSS model uses measurements along the aircraft flight track as input, the model is more sensitive to point-by-point variations in these parameters than a large-scale diurnal model, and is therefore more appropriate to compare with flight track data.

Background aerosol conditions

The values of NO_x/NO_y for the September flight are shown in Fig. 2 along with the LSS model calculations. In the segment from 42° N to 48° N, the aerosol SA values are identified as background values from comparisons with previous aerosol and condensation nuclei measurements^{9,27}. Some volcanic material may be present, however, without significant increases in SA. The LSS results using only gas-phase reactions are higher than the present observations at all latitudes. The inclusion of reaction (1), with a reaction efficiency⁸ of 0.1 and a rate proportional to observed aerosol SA, reduces NO_x/NO_y by nearly a factor of two, providing much better agreement with the observations. For background aerosol, the heterogeneous model calculations are lower than observations. The agreement would be further improved by lowering the reaction probability



for reaction (1) by a factor of two, which is the recommended uncertainty limit⁸. Equally important in this model comparison are the relative changes in NO_x/NO_y near 40° N where aerosol increases above background values. Here, the gas-phase model results show no features comparable to the observations while the absolute and relative changes are well presented in the heterogeneous case. Thus, the data in Fig. 2 alone provide strong qualitative and quantitative evidence for the need to include reaction (1) in the evaluation of photochemistry in the lower stratosphere. In a separate calculation, reaction (2) does not further reduce NO_x/NO_y for the limited range of temperature and aerosol SA values in Fig. 2.

The flight segment at background aerosol values can also be used to demonstrate the sensitivity of the LSS calculations to OH, the key unmeasured species in these calculations. In Table 1, greater sensitivity is found for NO_x/NO_y in the gas-phase case because reaction with OH is the only pathway for NO_x conversion to HNO_3 . In the heterogeneous case, the sensitivity is reduced because the added reactive pathway of reaction (1) not only reduces NO_x , but also competes with the OH reaction. *In situ* measurements of OH have been obtained at mid-latitudes for altitudes between 23 and 38 km (ref. 29). Values calculated

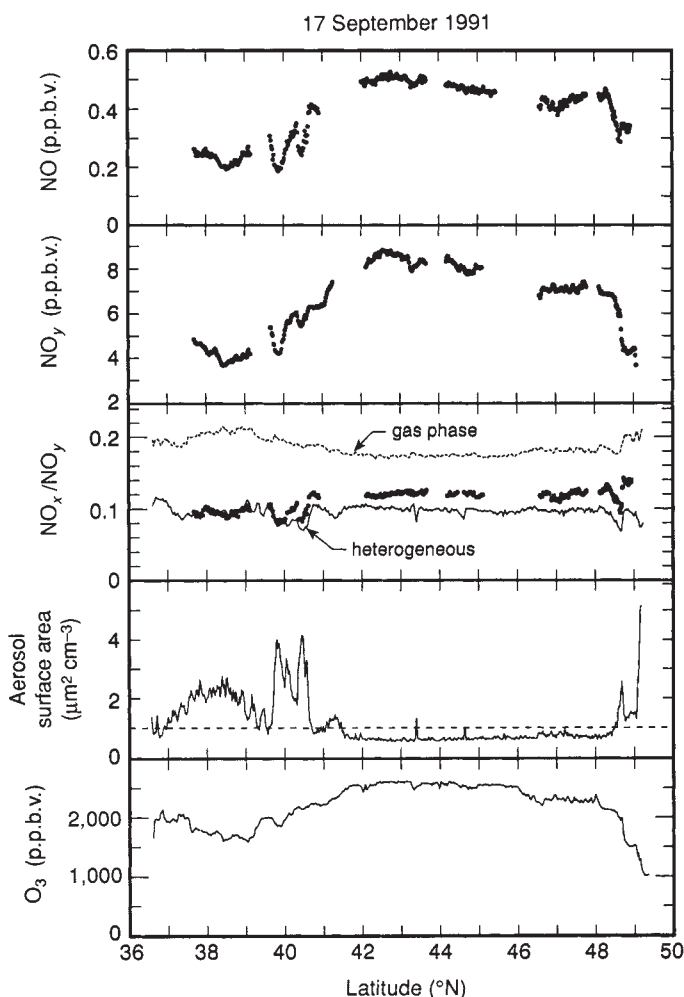
with the diurnal model used above agree with the OH observations to better than 20% over most of the altitude range, but observations exceed model values by 50% near 23 km. In contrast, lower OH may be expected to result from the scavenging of OH in reaction with sulphate aerosol^{30,31}. If, in future measurements, OH values are found to be substantially higher than calculated with current chemistry, then the relative importance of reaction (1) may be reduced significantly at background aerosol levels. However, the conclusion that NO_x has a diminished role in the contemporary atmosphere will remain unchanged.

The $\text{N}_2\text{O}_5 + \text{H}_2\text{O}$ heterogeneous reaction has significant implications for ozone loss calculations. In both ground-based and satellite observations^{32,33}, downward trends in ozone abundance are found at mid-latitudes that are larger than calculated in standard gas-phase models. Calculations of ozone depletion rates that include reaction (1) show, first, that an increase in aerosol SA leads to a decrease in NO_x/NO_y and the contribution of the NO_x catalytic loss cycle; second, an increase in the contribution of Cl_y - and HO_x -catalysed loss of ozone; and third, approximately no change in the total loss rate of ozone in the contemporary atmosphere for aerosol surface areas up to

FIG. 2 Data from the flight of 17 September 1991 as a function of latitude. The flight started at Moffett Field, California (37° N, 123° W) and proceeded north. Individual points are 10-s-average values corresponding to 2 km distance along the flight track. Data gaps are due to calibration procedures. Data for aerosol SA represent a sum over particle diameters of 0.07 to 3 μm obtained using a laser aerosol spectrometer^{22,27}. Surface area for diameters > 3 μm is negligible. The surface mean diameter varied from 0.35 μm for background conditions (surface area < 1 $\mu\text{m}^2 \text{cm}^{-3}$) to near 1 μm for enhanced conditions. Measured condensation nuclei values are 5–7 cm^{-3} during background conditions²⁷. The aerosol H_2SO_4 weight per cent is calculated to be near 70. The dashed line represents twice background aerosol surface area (1.0 $\mu\text{m}^2 \text{cm}^{-3}$).⁹ The range of other parameters are: pressure: 50–60 m bar; altitude: 19.5–20.5 km; temperature: 213–217 K; potential temperature: 480–510 K; SZA: 38°–50°; ClO : 0.02–0.04 parts per billion by volume (p.p.b.v.); water vapour: 4–5 parts per million by volume (p.p.m.v.) (K. Kelly, personal communication). In the NO_y instrument, component species are converted to NO on a gold catalyst with subsequent detection by the chemiluminescent reaction with reagent ozone⁴⁷. A separate chemiluminescence channel detects ambient NO. *In situ* measurements of ClO and O_3 use resonance fluorescence and ultraviolet absorption, respectively^{48,49}. Aerosol particles are measured with two spectrometers in flight with good agreement²². NO_x is estimated from the ER-2 observations using the expression²⁸

$$[\text{NO}_x] = [\text{NO}] (1 + (k_a[\text{O}_3] + k_b[\text{ClO}]) / J_{\text{NO}_2}) \quad (3)$$

where K_a and K_b are the reaction rates of NO with O_3 and ClO , respectively, to form NO_2 , as calculated using measured temperature and pressure⁸. J_{NO_2} , the photolysis rate of NO_2 in the sampled air parcel calculated from a radiation scattering model, has a near constant value of 0.014 s^{-1} in both data sets²⁶. In the lower stratosphere at mid- to high latitudes, the reactions of NO_2 with O and of NO with HO_2 can be neglected in formulating reaction (3)⁴⁰. The NO_2/NO ratio in this segment is near unity. OH is calculated separately along the flight track for the gas phase and heterogeneous cases. In the NO_x/NO_y panel, the lines represent the LSS model calculations for the gas-phase case (dashed) and the heterogeneous case with reaction (1) (solid). The aircraft observations and calculations of photolysis rates and OH are incorporated directly into the LSS calculations at each point along the flight track. The ratios NO_2/NO and $\text{ClONO}_2/\text{NO}_x$ are calculated from a steady-state approximation²⁸ whereas $\text{N}_2\text{O}_5/\text{NO}_x$, HNO_3/NO_x and $\text{HO}_2\text{NO}_2/\text{NO}_x$ are calculated as diurnal averages, a valid assumption for measurements taken near solar noon as the calculated quantities reach their average values near noon. These ratios are then combined to give the NO_x/NO_y ratio. The LSS model makes use of recent studies of the temperature dependence of the HNO_3 photolysis cross-section^{50,51}. However, when these studies are combined to calculate J_{HNO_3} for the flight segment presented here, differences from the current recommended cross-sections⁸ are small (see Table 1) (J. B. Burkholder, R. A. Cox and



A. R. Ravishankara, personal communication). The uncertainty in NO_x/NO_y values calculated from the observations is estimated as $\pm 30\%$. This estimate includes uncertainties that are associated with the reaction rate constants and the photolysis rate in reaction (3) and other rates used in the model calculations²⁸. Recommended rate constants are used in all cases⁸. The overall uncertainty in the model comparisons is comparable to the difference between the observed and model NO_x/NO_y values.

background values^{4,19}. Although reaction (1) does not increase the rate of ozone loss as summed over known catalytic cycles, the results of ozone trend calculations for the period 1979–90 that include reaction (1) agree better with observed loss over this period^{19,33}. This is a result of the increased importance of the chlorine- and bromine-based loss and the known growth in anthropogenic chlorine and bromine over this decade. Thus, heterogeneous chemistry is a plausible explanation for the model discrepancy in downward trends of ozone at mid-latitudes¹⁹.

Aircraft emissions are a sizeable source of NO_x in the upper troposphere and lower stratosphere and consequently can affect photochemical ozone production and loss in both regions^{3,34–36}. The NO_x emitted by a newly proposed fleet of supersonic aircraft is expected to be comparable to natural NO_x production in the stratosphere. Calculations of ozone loss in current gas phase, two-dimensional photochemical models show significant reductions in ozone as a result of NO_x emissions from the proposed fleet^{3,34}. These calculations are invalid because gas-phase models do not represent atmospheric observations of reactive nitrogen as demonstrated in Fig. 2. With the inclusion of reaction (1), the models not only agree better with observations but hemispheric ozone loss is reduced to negligible

levels for typical schemes of fleet usage. Ozone loss is reduced because the NO_x catalytic loss cycle is slowed as NO_x is converted to less reactive HNO_3 (ref. 3). Thus, reaction (1) effectively lowers ozone sensitivity to NO_x emissions, while increasing that of chlorine emissions as noted above. However, other issues remain concerning NO_x emissions, such as the indirect effect on the formation of PSCs and the associated release of reactive chlorine in polar regions³⁷.

Volcanic aerosol conditions

The aerosol SA values in Fig. 3 show the striking effect of Mount Pinatubo. Background conditions as in Fig. 2 were infrequently observed in September and completely absent in February–March in the lower stratosphere. The September and March flights have unprecedented potential for evaluating the role of reactions (1) and (2) in the atmosphere. First, the large SA differences are expected to enhance the detection of any related photochemical changes. The uniformity of enhanced levels of aerosol SA found throughout the later flight (22° N–45° N) suggests that observed or derived changes will probably be representative of changes throughout the hemisphere at mid-latitudes in the lower stratosphere. Second, each flight occurs within days of fall or spring equinox at similar latitudes, resulting

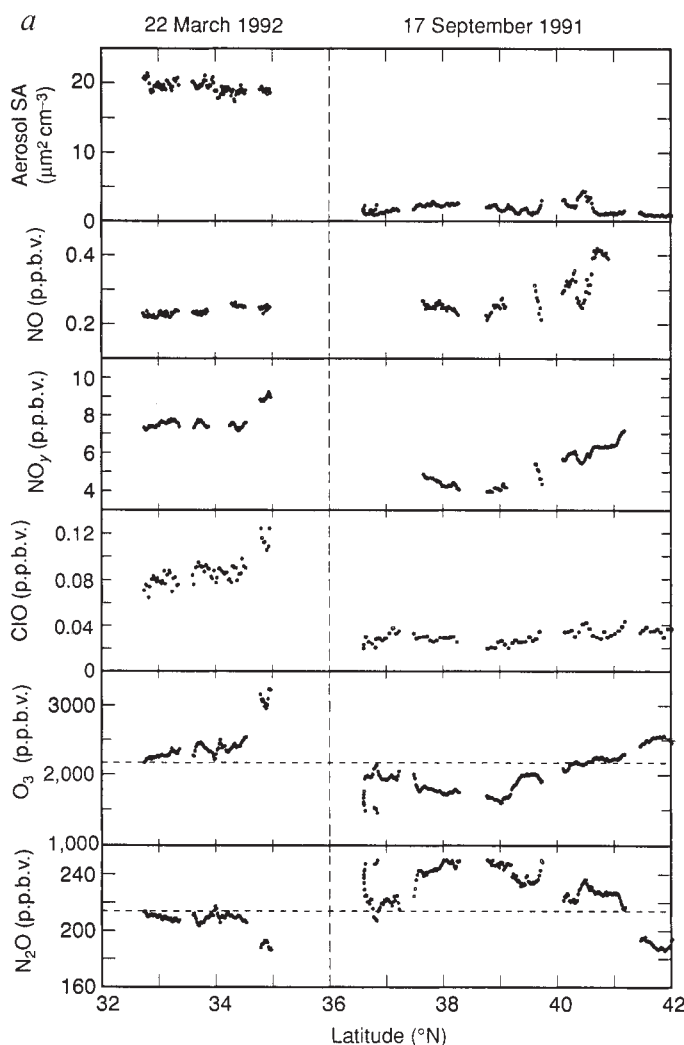
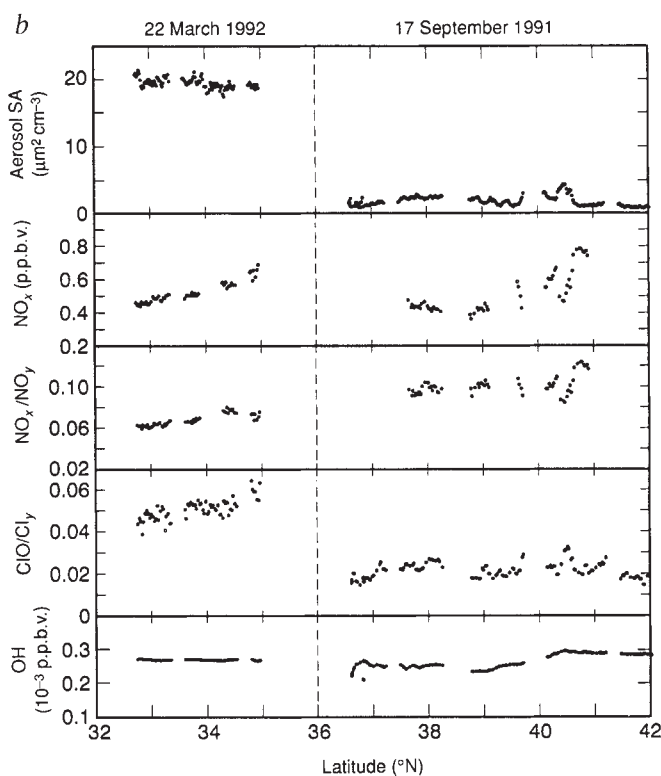


FIG. 3a, Measured species for the September and March flights where the selection criteria of $180 < \text{N}_2\text{O} < 250$ p.p.b.v. and $35^\circ < \text{SZA} < 40^\circ$ (after solar noon) have been applied. The March flight proceeded south from Bangor, Maine (69° W). Individual points are 10-s averages corresponding to 2 km spatial resolution. The vertical dashed line separates data from the two flights. The combined range of meteorological parameters is: temperature, 210–220 K; pressure, 50



– 65 mbar; altitude, 19–20.5 km; potential temperature, 460–510 K. Surface mean diameters are near $1.1 \mu\text{m}$ in March. The uncertainty in ClO for the September data ($\pm 60\%$) is higher than for the March data ($\pm 30\%$). All measurements were made within 2 hours after solar noon. For equinox conditions here, the observation latitude is the approximate minimum diurnal SZA. The dashed lines in the O_3 and N_2O panels serve as a guide to evaluate changes in O_3 at constant N_2O between the September and March flights⁶. *b*, Derived values of NO_x , NO_y/NO_y and ClO/Cl_y using the measurements in *a*. The inorganic chlorine reservoir, Cl_y , is estimated as the difference between total available chlorine (3.4 p.p.b.v.) and organic chlorine, derived using the empirical relation with measured N_2O (ref. 22). More precise values of total chlorine that result from age-of-air corrections would alter Cl_y only by ± 0.2 p.p.b.v. and hence are not important here. For these low ClO values, $(\text{ClO})_2$ is negligible.

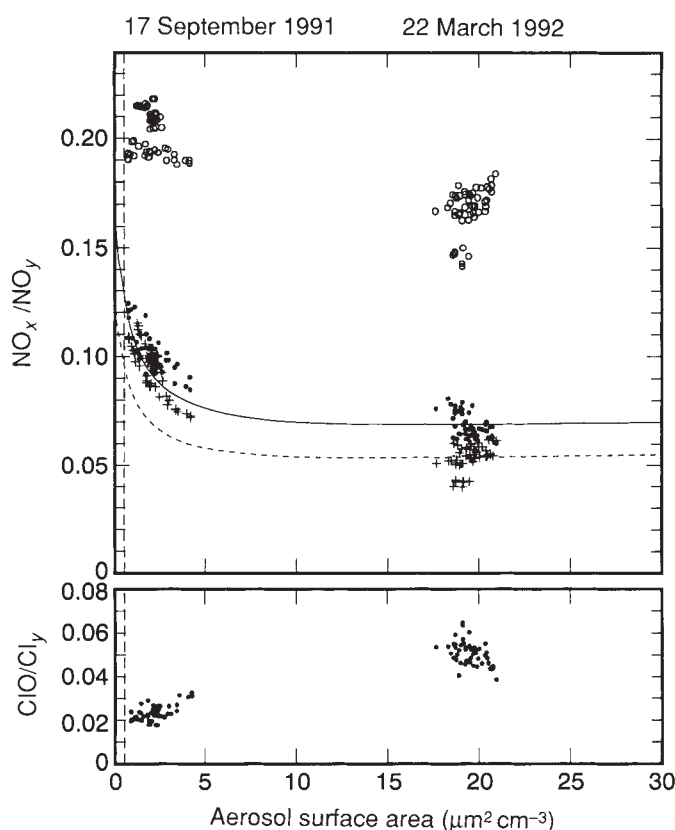


FIG. 4 Scatter plot of the NO_x/NO_y and ClO/Cl_y data from Fig. 3b with observed aerosol SA (solid circles). Gas phase (open circles) and heterogeneous case model calculations (crosses) are included, where the latter use observed aerosol SA with reaction (1). The SA scale for the gas phase case has no meaning except to separate data from the two flights. The vertical dashed line represents background aerosol surface area ($0.5 \mu\text{m}^2 \text{cm}^{-3}$)⁹ OH values from Fig. 3b are used in the calculations for both model cases. The curved lines represent the dependence on aerosol SA in the model heterogeneous case for the average conditions in the September (solid) and March (dashed) data sets.

in comparable diurnal variations of solar illumination in the air parcels for each flight. Third, isentropic back trajectories for both flights show that recent latitude excursions and associated photochemical changes are small (less than $\pm 10^\circ$) implying that observations are representative of the latitude at which the air parcels are sampled. Fourth, similar, low SZAs occur in each flight. When SZAs are similar with respect to solar noon, NO_x/NO_y and ClO/Cl_y ratios can be directly compared between flights near equinox. For low SZAs ($< 60^\circ$), the scattering of solar radiation due to enhanced aerosol SA can be neglected in the calculation of photolysis rates³⁸. Also, near solar noon, photolysis of reservoir species produce NO and ClO amounts well above instrumental detection limits.

Figure 3a displays the trace species measurements from both flights where all data meet the criteria of $180 < \text{N}_2\text{O} < 250$ parts per billion by volume (p.p.b.v.) and $35^\circ < \text{SZA} < 40^\circ$. The N_2O constraint provides a limited altitude range in the lower stratosphere³⁹. The SZA constraint is chosen to be as small as possible while providing a large data set at similar latitudes that includes the observed range of aerosol SA values. Figure 3b shows values derived for NO_x , NO_x/NO_y and ClO/Cl_y . Between the September and March flights, the NO_x/NO_y ratio shows some decrease while a large increase occurs in ClO/Cl_y . To highlight the observed changes, the ratios are shown in Fig. 4 against SA. In the early flight, NO_x/NO_y decreases significantly as SA increases up to 10 times background values, displaying some evidence of nonlinearity. These changes are also evident in Fig. 2. ClO/Cl_y in this flight increases with increasing aerosol

SA, but with less evidence of non-linearity. With the March flight, a further decrease in NO_x/NO_y is observed along with a larger increase in ClO/Cl_y as aerosol values increase to 40 times background values. The combined data for NO_x/NO_y suggest that the decrease in NO_x/NO_y is nonlinear over the observed aerosol SA range. With no additional qualification, the two data sets serve to limit changes in NO_x/NO_y and ClO/Cl_y that can be expected at mid-latitudes for the observed SA increases. The changes fall far short of the near-complete removal of NO_x necessary to cause catastrophic ozone loss rates⁴.

The role of reaction (1) in the change of NO_x/NO_y is demonstrated by comparison to the LSS model results. As in Fig. 2, the results for the gas-phase case shown in Fig. 4 are significantly higher than observations of NO_x/NO_y for both flights. Indeed, the failure of the gas-phase model to represent the observations is systematic throughout the AASE II data set at these and higher latitudes⁴⁰. The smaller differences in gas-phase NO_x/NO_y found between flights follow from the effect of changes in observed O_3 and ClO which serve as input to the LSS model. The results for the heterogeneous case including reaction (1) agree much better with the observations for both flights. The lifetime of N_2O_5 in reaction (1) is reduced from several days to less than one day for the aerosol SA increase in Fig. 4 (refs 5,9). For comparison, the HNO_3 lifetime is roughly 10 days at noon in this latitude and altitude range⁴¹. With the September data set, the model reproduces the non linear decrease of NO_x/NO_y with slightly smaller values. In both cases, the model underestimates the observations on average. These systematic differences are within the uncertainty of the comparison, but may result from reaction probabilities that are too high for ambient aerosol.

To provide a connection between the data sets, average measured values of the model input parameters for each flight were used as input into a separate model run in which aerosol SA alone was varied over the full range in Fig. 4. The resulting curves show that reaction (1) produces the expected nonlinear, saturation response in NO_x/NO_y . The offset between the curves is consistent with the changes in the gas-phase photochemical balance between flights. Thus, the two curves would be expected to overlay if the flights sampled air with fully identical photochemical conditions. Although the gas-phase model results in Fig. 4 are plotted against aerosol SA for clarity, the average values for each flight could be plotted as the intercepts of the model curves (SA=0), emphasizing the nonlinear NO_x/NO_y response. With varying quantitative features, a similar saturation response is calculated throughout the vertical extent of the stratospheric aerosol layer¹⁶. The NO_x/NO_y response in the model saturates in the aerosol SA range of 5–10 $\mu\text{m}^2 \text{cm}^{-3}$ in Fig. 4. The data do not yield a more precise saturation value. A lower heterogeneous rate for reaction (1), as suggested by the differences between model and observations at each point, would increase the apparent saturation value in the model.

Including reaction (2) for the observed temperature range of 210–220 K does not change the model results for the September data, but further lowers the March NO_x/NO_y results by 0.01–0.02. Thus, reaction (2) plays a minor role in the decrease of NO_x/NO_y under these conditions. A higher reaction probability usually associated with lower temperatures is necessary to further reduce NO_x/NO_y and approach catastrophic ozone loss rates⁴. A model simulation of the increase in ClO/Cl_y in Fig. 4 may show that reaction (2) plays a more important role within the Cl_y reservoir. In addition, in higher-latitude regions where HNO_3 photolysis rates and ambient temperatures are lower, reaction (2) will compete more favourably with reaction (1) in altering NO, and Cl_y partitioning^{42,43}.

From Fig. 4, a negative correlation exists between NO_x/NO_y and ClO/Cl_y . A different LSS model configuration incorporating NO_x/NO_y as an input parameter is needed to predict changes in ClO within the Cl_y reservoir. However, other model studies

have concluded that observed ClO values with background aerosol are higher than expected from gas-phase models and that better agreement is found when reaction (1) is included^{10–12}. A ClO increase with respect to Cl_y follows from decreases in the formation rates of HCl and ClONO₂ as NO_x is reduced. Although the changes in ClO are small with respect to the inactive reservoir sum (HCl + ClONO₂), significantly larger increases in ClO/Cl_y would be expected if the reduction in NO_x/NO_y did not saturate with increasing aerosol SA⁴.

As a further consideration, we note that between the September and March flights, processing of stratospheric air by reactions on PSCs occurred in the northern polar vortex, liberating chlorine from the HCl and ClONO₂ reservoirs and converting NO_x to HNO₃ (refs 1,2; and C. Webster *et al.*, manuscript in preparation). Although subsequent transport and mixing of this processed air to lower latitudes probably occurs, quantitative features are difficult to evaluate⁴⁴. After transport, the long time constant for the reformation of HCl could result in elevated ClO values in the mid-latitude stratosphere in early spring^{44,45}. However, the formation of PSCs ended in early to mid February in 1992, allowing substantial time for HCl recovery (C. Webster *et al.*, manuscript in preparation). In any case, the observed increases in ClO in March are an upper limit for changes induced by increased aerosol SA. The effect on the NO_y reservoir is likely smaller because of a shorter lifetime for NO_x. Because NO_x/NO_y is reduced by PSC processing, transport of polar air to mid-latitudes does not invalidate the

conclusion of a non linear change in NO_x/NO_y with aerosol SA, although the apparent value of aerosol SA at saturation may be changed somewhat.

On the basis of the combined evidence from the data and model simulations presented here, changes in the NO_x/NO_y ratio which saturate well below aerosol SA values of 20 μm² cm⁻³ can be explained by the effect of reaction (1) on the partitioning of the NO_y reservoir. The proportionately larger changes in the ClO/Cl_y ratio are consistent with the role of ClONO₂ and HCl in the NO_y and Cl_y reservoirs. Reaction (2) plays a minor role in affecting the NO_x/NO_y ratio in these observations. The consequences for ozone depletion from the effect of volcanic aerosol on NO_x and ClO are the subject of other studies^{3,4,19,23,24,33}. However, catastrophic loss rates are precluded for the observation conditions because the saturation of the NO_x/NO_y response limits the increase in ClO/Cl_y. Indeed, using the reference correlation between observed O₃ and the long-lived tracer N₂O (ref. 47), the data in Fig. 3 show that large changes in ozone did not occur between September and March for air parcels of similar photochemical history. The possibility of smaller changes in loss rates depends on details of the compensating effects of reduced loss from the NO_x catalytic cycle and an increased loss from ClO and HO_x cycles^{4,5}. The data presented here will aid in the proper evaluation of these losses when used to constrain trace constituents in photochemical transport models of the atmosphere (R. J. Salawitch *et al.*, manuscript in preparation). □

Received 21 January; accepted 4 May 1993.

- Anderson, J. G., Toohey, D. W. & Brune, W. H. *Science* **251**, 39–52 (1991).
- Solomon, S. *Rev. Geophys.* **26**, 131–148 (1988).
- Weissenstein, D. K. *et al. Geophys. Res. Lett.* **1**, 1991–1994 (1991).
- Prather, M. J. *Geophys. Res.* **97**, 10187–10191 (1992).
- McElroy, M. B., Salawitch, R. J. & Minschwaner, K. *Planet. Space Sci.* **40**, 373–401 (1992).
- Hanson, D. R. & Ravishankara, A. R. *J. Geophys. Res.* **96**, 17307–17314 (1991).
- Van Doren, J. M. *et al. J. Phys. Chem.* **95**, 1684–1689 (1991).
- Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling, Evaluation No. 10, JPL Pub. 92-20* (Jet Propulsion Laboratory, NASA Administration, 1992).
- Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5041 (1989).
- Mather, J. H. & Brune, W. H. *Geophys. Res. Lett.* **17**, 1283–1286 (1990).
- King, J. C. *et al. Geophys. Res. Lett.* **18**, 2273–2276 (1991).
- Avallone, L. M. *et al. Geophys. Res. Lett.* (submitted).
- Wofsy, S. C. *J. geophys. Res.* **83**, 364–378 (1978).
- Cadle, R. D. *et al. J. Geophys. Res.* **80**, 3381–3385 (1975).
- Considine, D. B. *et al. Geophys. Res. Lett.* **19**, 397–400 (1992).
- Mills, M. J. *et al. Geophys. Res. Lett.* (in the press).
- Russell, J. M. III *et al. J. Geophys. Res.* **93**, 1718–1736 (1990).
- Keys, J. G. *et al. Nature* **361**, 49–51 (1993).
- Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Nature* **352**, 134–137 (1991).
- McCormick, M. P. & Veiga, R. E. *Geophys. Res. Lett.* **19**, 155–158 (1992).
- McCormick, M. P. & Trepte, C. R. *J. Geophys. Res.* **92**, 4297–4306 (1987).
- Wilson, J. C. *et al. Geophys. Res. Lett.* (submitted).
- Brasseur, G. & Granier, C. *Science* **257**, 1239–1242 (1992).
- Granier, C. & Brasseur, G. *J. Geophys. Res.* **97**, 18015–18033 (1992).
- Logan, J. A. *et al. Phil. Trans. R. Soc. A* **290**, 187–234 (1978).
- Anderson Jr., D. E. & Lloyd, S. A. *J. Geophys. Res.* **95**, 7429–7434 (1990).
- Wilson, J. C. *et al. J. Geophys. Res.* **97**, 7997–8013 (1992).
- Kawa, S. R. *et al. J. Geophys. Res.* **97**, 7905–7923 (1992).
- Stimpfle, R. M. *et al. J. Geophys. Res. Lett.* **17**, 1905–1908 (1990).

- Hanson, D. R. *et al. J. Phys. Chem.* **96**, 4979–4985 (1992).
- Gershenzon, Y. M. *et al. Kinet. Katal.* **27**, 923 (1987).
- Stolarski, R. *et al. Science* **256**, 342–349 (1992).
- WMO/UNEP *Global Ozone Research and Monitoring Project Rep. No. 25*. (World Meteorological Organization, Geneva, 1991).
- Johnston, H. S., Kinnison, D. E. & Wuebbles, D. J. *J. geophys. Res.* **94**, 16351–16363 (1989).
- Johnson, C., Henshaw, J. & McInnes, G. *Nature* **355**, 69–71 (1992).
- Ehhalt, D. H., Rohrer, F. & Wahner, A. *J. geophys. Res.* **97**, 3725–3738 (1992).
- Bruhl, P. C. & Crutzen, P. J. *Geophys. Res. Lett.* **18**, 1465–1468 (1991).
- Anderson, D. E. & Demajistre, R. *Eos* **73**, 112 (1992).
- Loewenstein, M. *et al. J. geophys. Res.* **94**, 11589–11598 (1989).
- Kawa, S. R. *et al. Geophys. Res. Lett.* (submitted).
- Kawa, S. R. *et al. J. geophys. Res.* **95**, 18597–18609 (1990).
- Solomon, S. *et al. Nature* **363**, 245–248 (1993).
- Cox, R. A. *et al. Geophys. Res. Lett.* (submitted).
- Tuck, A. F. *et al. J. geophys. Res.* **97**, 7883–7904 (1992).
- Brune, W. H. *et al. Science* **252**, 1260–1266 (1991).
- Proffitt, M. H. *et al. Nature* **347**, 31–36 (1990).
- Fahey, D. W. *et al. J. geophys. Res.* **94**, 11299–11315 (1989).
- Brune, W. H. *et al. J. geophys. Res.* **94**, 16649–16663 (1989).
- Proffitt, M. H. *et al. J. geophys. Res.* **94**, 16547–16555 (1989).
- Rattigan, O. *et al. J. Photochem. Photobiol. A: Chem.* **66**, 313–326 (1992).
- Burkholder, J. B. *et al. J. geophys. Res.* (submitted).

ACKNOWLEDGEMENTS. We thank J. G. Anderson, M. Prather, J. M. Rodriguez, S. Solomon and A. F. Tuck for discussions, and J. Burkholder, R. K. Talukdar, A. R. Ravishankara, and K. K. Kelly for making data available before publication. We acknowledge the many individuals who made the 1991–92 Airborne Arctic Ozone Expedition II a success, especially the pilots of the NASA ER-2 aircraft. This research is supported in part by the Upper Atmosphere Research Program of NASA and the Atmospheric Effects of Stratospheric Aircraft component of the NASA High Speed Research Program.

Alteration in a new gene encoding a putative membrane-organizing protein causes neuro-fibromatosis type 2

Guy A. Rouleau*, Philippe Merel†, Mohini Lutchman*, Marc Sanson*†, Jessica Zucman†, Claude Marineau*, Khé Hoang-Xuan†, Suzanne Demczuk*, Chantal Desmaze†, Béatrice Plougastel†, Stefan M. Pulst‡, Gilbert Lenoir§, Emilia Bijlsma||, Raimund Fashold¶, Jan Dumanski#, Pieter de Jong**, Dilys Parry††, Roswell Eldridge‡‡, Alain Aurias†, Olivier Delattre† & Gilles Thomas†§§

* Centre de recherche en Neurosciences, McGill University, and Montreal General Hospital Research Institute, 1650 Cedar Avenue, Montreal, Quebec H3G 1A4, Canada

† Laboratoire de Génétique des Tumeurs, C/JF INSERM 9201, Institut Curie, 26 rue d'Ulm, 75231 Paris Cedex 05, France

‡ Neurogenetics Laboratory, Cedar-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA

§ International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon, Cedex 08, France

|| Institute for Human Genetics, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

¶ Institut für Humangenetik, Universität Erlangen-Nürnberg, Schwabachanlage 10, 8520 Erlangen, Germany

Department of Clinical Genetics, Karolinska Institutet PO Box 60500, 104 01 Stockholm, Sweden

** Human Genome Center, Lawrence Livermore National Laboratory, PO Box 5507, Livermore, California 94550, USA

†† Clinical Epidemiology Branch, National Cancer Institute, National Institute of Health, Bethesda, Maryland 20894, USA

‡‡ Retired, United States Public Health Service, Bethesda, Maryland 20894, USA

Neurofibromatosis type 2 (NF2) is a monogenic dominantly inherited disease predisposing carriers to develop nervous system tumours. To identify the genetic defect, the region between two flanking polymorphic markers on chromosome 22 was cloned and several genes identified. One is the site of germ-line mutations in NF2 patients and of somatic mutations in NF2-related tumours. Its deduced product has homology with proteins at the plasma membrane and cytoskeleton interface, a previously unknown site of action of tumour suppressor genes in humans.

THE neurofibromatoses are autosomal dominant diseases which primarily affect the nervous system. Neurofibromatosis type 1 (NF1), which occurs with an incidence of 1/3,000, predisposes mainly to the development of peripheral neurofibromas, café au lait macules, optic nerve gliomas and bony abnormalities. Neurofibromatosis type 2 (NF2), which occurs with an incidence of 1/37,000 (ref. 1), is mainly associated with the development of schwannomas, and also, to a lesser extent, to meningiomas and ependymomas². Juvenile cataracts are also frequently seen in NF2. Though clinical studies first suggested the existence of at least two different forms of neurofibromatosis, definite proof of the existence of two distinct syndromes was provided only recently when it was determined that the genetic defects causing NF1 and NF2 map to separate chromosomes. Linkage studies showed that, in contrast to NF1 which maps to chromosome 17

(refs 3,4), NF2 maps to chromosome 22 (ref. 5). Further genetic studies of NF2 have shown it to be a genetically homogeneous disease⁶. Deletion studies in tumours suggests that loss of the NF2 gene is also important in the development of sporadic schwannomas^{7,8} and meningiomas⁹⁻¹¹, which together represent 30% of all primary brain tumours. This implies that NF2 might be caused by mutations in a gene with tumour suppressor activity.

Segregation studies in affected families, and characterization of a germ-line deletion in one NF2 pedigree have determined that NF2 maps to the interval between D22S212 and D22S32 (ref. 12 and M. S. *et al.*, unpublished results), a region small enough to attempt the identification of the NF2 gene using a positional cloning strategy.

Cloning of the D22S212/D22S32 region

Using an extended panel of somatic cell hybrids, D22S212, D22S32 and four additional loci (Cos5/6, Kil764, Kil045, LIF) were sublocalized to two adjacent subregions (groups 7 and 8) that flank a translocation breakpoint recurrently observed in Ewing's sarcoma (refs. 13,14 and unpublished results). These loci were progressively expanded by the isolation of overlapping cosmids using chromosome 22-specific libraries¹⁵. In this process several contiguous sequences (contigs) were connected so that a total of four cosmid contigs was finally obtained. Isolation of yeast artificial chromosomes and/or Southern analysis of total human DNA provided an estimate of the sizes of the gaps between these contigs indicating that the distance between D22S212 and D22S32 was about one million base pairs (Fig. 1).

Search for submicroscopic chromosome alteration

Forty-two unrelated NF2 patients were screened for rearrangements in the D22S212/D22S32 region using pulsed-field gel

TABLE 1 Rearrangements in NF2 patients detected by N1.1 cDNA

Patient	Restriction enzyme	Size (kb)	
		Normal	Abnormal
4774	<i>SfiI</i>	50	70
	<i>BssHII</i>	130	210
	<i>EagI</i>	115	195
R2642	<i>BssHII</i>	130	145
	<i>EagI</i>	115	130
4702	<i>EcoRI</i>	20	19
	<i>HindIII</i>	7	6
	<i>HincII</i>	13	12
R267	<i>EcoRI</i>	20	19

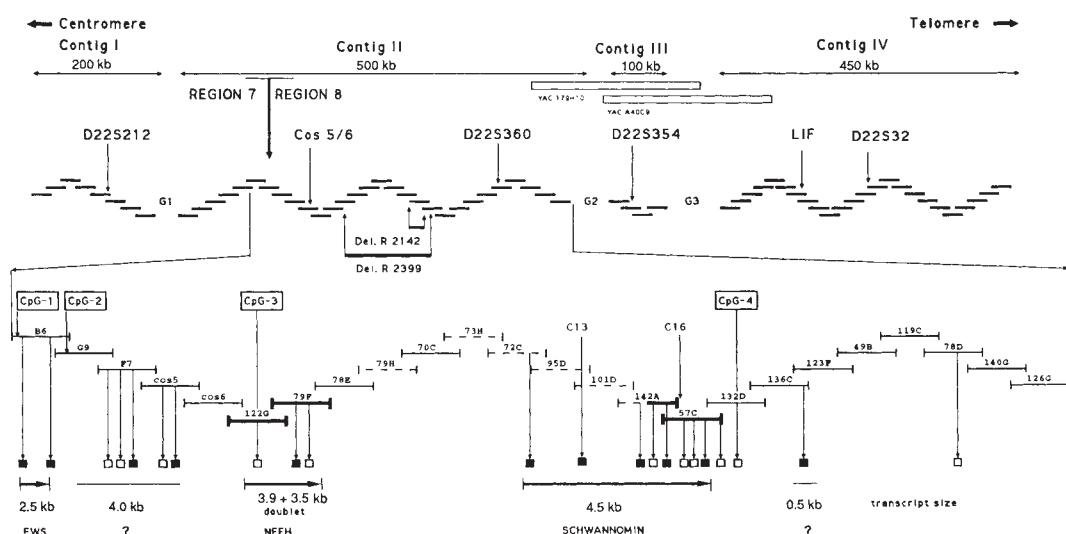
None of the abnormal bands were present in 40 control individuals tested by PFGE (individuals R2642 and 4774) and 90 control individuals by the Southern blots.

TABLE 2 Oligonucleotides used to screen for mutation in the SCH gene

	Region screened (codon number)	Oligonucleotide sequence
Set 1	39 to 80	5' GC clamp 1- TTGCTCACAGTGTCTTCCC 3' 5' M13 reverse- TCAGCCCCACCAGTTTCATC 3'
Set 2	150 to 172	5' M13 direct- ATCTTTAGAATCTCAATCGC 3' 5' GC clamp 2- AGCTTTCTTTAGACCACAT 3'
Set 3	226 to 270	5' M13 direct- CCACAGAATAAAAAGGGCAC 3' 5' GC clamp 2- GATCTGCTGGACCCATCTGC 3'
Set 4	334 to 374	5' M13 direct- TCGAGCCCTGTGATTCAATG 3' 5' GC clamp 2- AAGTCCCCAAGTAGCCTCCT 3'
Set 5	376 to 446	5' GC clamp 1- CCCACTTCAGCTAAGAGCAC 3' 5' M13 reverse- CTCCTCGCCAGTCTGGTG 3'
Set 6	527 to 579	5' M13 direct- TCTCACTGTCTGCCCAAG 3' 5' GC clamp 2- GATCAGCAAAATACAAGAAA 3'
5' GC clamp 1 = 5' CCCC GCCCGGCCGCGCCCGCCCCCGCCCCCTCCCGGCCCGCCCCCTGGCGCCCCGC 3'		
5' GC clamp 2 = 5' CCCCACGCCACCCGACGCCCGCCGACCCCGCCCGCGCCCGGCCCGCCCCGC 3'		
5' M13 direct = 5' TGTAAACGACGGCCAGT 3'		
5' M13 reverse = 5' CAGGAACAGCTATGACC 3'		

Six genomic sequences containing in each case a whole exon were analysed using the computer programs Meltmap and 5QHTX³⁵ to determine the optimal set of two oligonucleotides. One member of each pair of oligonucleotides contains a GC-rich sequence (GC clamp 1 or 2)²⁸. The other member contains a sequence (M13 direct or M13 reverse) used to sequence the amplified product with commercially available fluorescent primers. All PCR reactions were done in 10 µl with AmpTaq Kit (Cetus) using a Perkin Elmer 9600 DNA cyclor (35 cycles with denaturing step 95 °C for 30 s, annealing temperature as specifically indicated in each case for 30 s and elongation at 72 °C for 90 s). Denaturing gradient gel electrophoresis (DGGE) was done as published⁴⁴. Specific conditions for PCR amplification and DGGE analysis were as follows: PCR annealing temperature, set 1 and 4, 62 °C; set 2, 3, 5 and 6, 60 °C. DGGE analysis: denaturation gradients (100% denaturant is 7M urea and 40% formamide), set 1, 2, 3 and 6, 30% to 80%; set 4 and 5, 40% to 90%. Electrophoresis was done for set 1, 3, 4 and 6 for 3 h; set 2 for 2 h and set 5 for 4 h. Amplified products were purified on Centricon 100 ultra-filtration devices (Amicon, Epernon, France)²² and direct sequencing was done using PRISM Taq fluorescent primer Idt (Applied Biosystems).

FIG. 1 Regional map of the NF2 locus. The relative position of the four contigs of cosmids, their size and their orientation with respect to the centromere and telomere of the long arm of chromosome 22 are shown. The minimum number of cosmids (tiling path, each bar corresponds to one cosmid) and the loci that were initially expanded are indicated below. The boundary between region 7 and 8 defined by a chromosome 22 mapping panel of somatic cell hybrids is indicated¹³. G1 to G3 refer to the three gaps between the four contigs. G1 has been estimated to be about 0.5 kb by Southern blotting experiments. G2 and G3 are about 100 kb and 20 kb, respectively, and have each been spanned with a YAC clone (M. Giovannini & L. Selleri, unpublished results). The positions of the deletion detected in individuals R2142 and R2399 are shown. In the lower part of the figure a portion of the tiling path of contig II is shown in greater detail. Cosmids that have been shown by FISH to be preserved or deleted in the NF2 patient R2399 are shown in thick or dotted lines, respectively (cosmid 142A is split by the deletion). The position of four clusters of GC nucleotides are indicated as follows: CpG-1 contains site(s) for *MluI*, *Bss*HII and *Sac*II; CpG-2, CpG-3 and CpG-4 each contain site(s) for *Sac*II, *Bss*HII and *Not*I. *Eco*RI fragments demonstrating phylogenetic conservation with rodent DNA (rat and mouse) are indicated by squares. They are black when the fragment is able to detect a transcript in RNAs from at least one of the tested human cell lines. The size of the detected transcript is indicated. When known, the direction of transcription is indicated by an arrow. The positions of the C13 and C16 probes are indicated.



METHODS. Expansion of six loci from the vicinity of the NF2 gene (D22S212, Cos5/6, Kii764, Kii045, LIF and D22S32) was obtained as previously described¹⁵. FISH was done as described in ref. 41. Each cosmid between B6 and 126G was digested with *Eco*RI. Aliquots were double digested with the following enzymes containing two CG dinucleotides in their target sequence (*Bss*HII, *Sac*II, *Not*I, *Nru*I and *Mlu*I) and alteration in the migration pattern was analysed. Individual *Eco*RI fragments were also gel purified and used as probes on Southern blots made from *Eco*RI digested human rat and mouse DNAs⁴². To reveal phylogenetic conservation two stringency conditions were used for washing (2 × SSC, 50 °C and 0.1 × SSC, 65 °C in both cases for 5 min). Northern blots were done using standard conditions⁴² with RNAs extracted from the following cell lines: HL60 (promyelocytic leukaemia), HeLa (cervix carcinoma), SKNBZ (neuroblastoma), 2102 EP and Tera 2 (teratocarcinomas), CCL225 (colon carcinoma), OZ (glioma), EW24 (Ewing's sarcoma), Bewo (choriocarcinoma).

electrophoresis (PFGE) and single-copy probes derived from the previously identified cosmids. Two individuals had abnormal fragments when probed with C16 isolated from the cosmid 57C (Fig. 2a): patient R2399 and R2142 displayed the expected 130 kilobase (kb) *Bss*HII fragment plus new fragments of 85 kb and 90 kb, respectively. With the same probe, both patients also had abnormal band size with *Sac*II and *Eag*I. In contrast, C13, a probe 40 kb centromeric to C16 and which detects the same *Bss*HII fragment in normal individuals, failed to detect the rearrangements in R2142 and R2399, suggesting that C13 maps to an area deleted in both patients.

The presence and extent of the deletions were tested using fluorescent *in situ* hybridization (FISH) on interphase nuclei or metaphase chromosomes using nine cosmids (Fig. 1). Both copies of chromosome 22 of patient R2399 were detected with the two most centromeric (122G, 79F) and the most telomeric (57C) cosmids. In contrast, five cosmids located between these two borders (79H, 73H, 72C, 95D and 101D) hybridized to a single copy of chromosome 22, confirming the presence of a deletion. For cosmid 142A, the second most telomeric cosmid tested, the recurrent observation of a strong signal on one copy of chromosome 22 and a weak signal on the other suggested that it straddled the distal end of the deletion. We concluded that patient R2399 has a germ-line 130 kb interstitial deletion flanked by cosmids 79F and 142A. For patient R2142, similar FISH

experiments demonstrated that only cosmid 101D was deleted (Fig 2b). Cosmid 95D, which partially overlaps with cosmid 101D, gave variable signal on one copy of chromosome 22, compatible with a partial deletion. Cosmids 57C, 142A and 72C were not deleted, suggesting that the deletion in R2142 is about 40 kb. To identify candidate genes for NF2 the segment of contig II into which these two deletions map was further studied.

Search for genes in the D22S212/D22S32 region

We made a systematic analysis of over 100 *Eco*RI fragments isolated from the 24 cosmids shown on the lower part of Fig. 1. To identify clusters of CG dinucleotides, which frequently mark the 5' region of genes¹⁶, each fragment was examined for the presence of restriction sites which contain two CGs in their recognition sequence. Four clusters of at least three different restriction sites were found. Each *Eco*RI fragment was also tested for its ability to hybridize rodent DNA and, when phylogenetic conservation was found, to detect transcripts on northern blots made with RNAs from various human cell lines. When the different data generated by this analysis were put together on the map, the position of four genes became evident. The most centromeric gene, associated with the CpG-1 cluster, marks the limit between subregions 7 and 8 (ref. 13). It has been called EWS because of its constant rearrangement in Ewing's sarcoma¹⁷. The closest identified gene on the telomeric side of

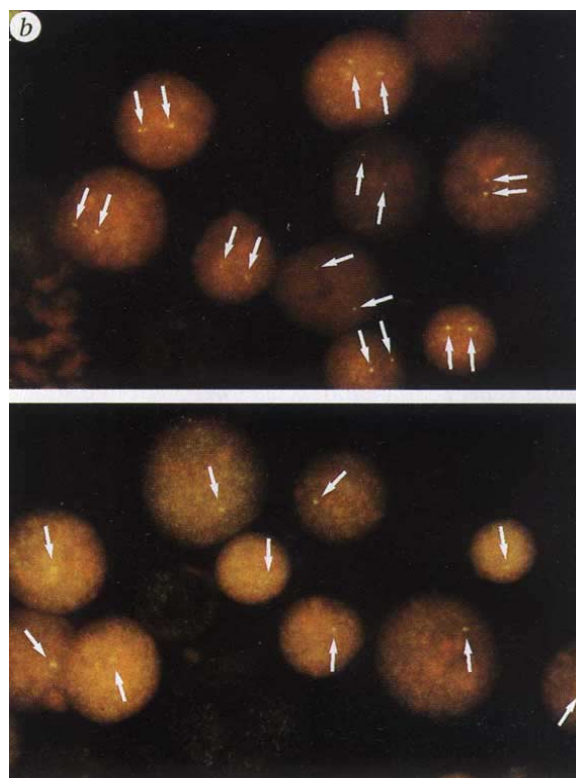
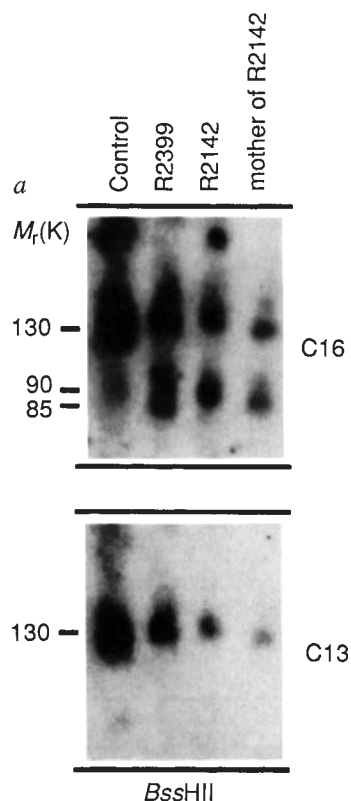
TABLE 3 germ-line and somatic SCH mutations in NF2-patients and NF2 related tumours

Sample	Mutation	Altered codon	Consequence	Family analysis
SP 10	CT deletion	54	Frame-shift	new mutation
SP 16	C to T	57	Non-sense	ND
IC 2	C to T	57	Non-sense	ND
GL 9	AGgt to AGtt	junction 80/81	Splice donor	segregates with NF2
ST 2	AGgt to AGtt	junction 80/81	Splice donor	ND
SP 4	agTA to atTA	junction 149/150	Splice acceptor	ND
EB D	GGgt to GGat	junction 172/173	Splice donor	segregates with NF2
EB N	C insertion	253	Frame-shift	new mutation
GL 18	C deletion	262	Frame-shift	ND
EB G	C to T	262	Non-sense	ND
GL 2C to T		341	ND	
SP 20	TGAACGC del	349 to 351	Frame-shift	new-mutation
RE 1	GAACGCAAGAGG replaced by CGAGAGAAGCA	350 to 353	Frame-shift	segregates with NF2
RF 10	T to C	360	Leu to Pro	segregates with NF2
GL 5	AGgt to AGat	junction 447/448	Splice donor	segregates with NF2
DNA from tumour				
IC-BE	C TO T*	57	Non-sense	Tumour type Chr.22 allele schwannoma preserved
28A	C to T*	262	Non-sense	schwannoma lost
IC-J	C to T†	262	Non-sense	schwannoma lost
IC-F	C to T†	341	Non-sense	schwannoma lost
IC-AB	GGAA del†	43 and 44	Frame-shift	meningioma preserved
IC-PO	TC del†	61	Frame-shift	meningioma lost

A group of 90 independent NF2 patients was screened for mutations on 6 different exons. Fifteen patients had an abnormally migrating band which was sequenced and revealed a mutation in each case. The predicted consequence of the mutation on the schwannomin structure is given. 'Segregates with NF2' means that the mutation was present in all studied affected members (at least two affected individuals in two different generations per family) and absent from all studied unaffected members. 'New mutation' indicates that the affected patients has two unaffected parents. In each of these three families, the SCH mutation of the patient was not present in either of his/her parents. The diagnostic criteria for NF2 were those defined by the 1991 NIH consensus conference statement³⁶. Non-paternity was excluded by typing the following highly polymorphic loci D5S346 and CAMBC (refs 37, 38). No familial relationship were found for patients GL 9 and ST 2 and for patients SP 16 and IC 2s ND. Other family members were not studied. Thirty schwannomas and 30 meningiomas from either NF2 or sporadic patients were also screened. Six tumours revealed the presence of an SCH mutation. DNAs from the blood of the six corresponding patients were also investigated. Schwannomas 28A and IC-BE developed in NF2 patients EB-G and IC 2, respectively. IC-J and IC-F correspond to the schwannomas of the sporadic patients J and F described in ref. 39. IC-AB and IC-PO are two meningiomas removed from sporadic patients. 'Lost', Somatic losses of heterozygosity in the tumour DNAs observed for the following loci: 28A: IGLV, BCR, CRYB2A, D22S32, D22S15, D22S29 and ARSA (ref. 40); IC-F: D22S170, IGLV, D22S32, PDGFB, D22S80 (ref. 25 and unpublished result); IC-J: D22S24, D22S9, D22S170 and D22S20 (ref. 39 and unpublished result); IC-PO: D22S20, PDGFB, D22S171 (unpublished result). 'Preserved', All observed chromosome 22 heterozygosities in the constitutional DNA were preserved in the tumour DNA.

*Mutation observed in the tumour is identical to that found in the patient blood DNA.

†Mutation is not observed in the patient blood DNA and is therefore somatic.



C

GGGTCTCTCGGGCCCATGCTGGCCGCTGGGGACCCGCGAGCCGACACCTTCCCGGGCGGCCACGCCACCATGGTGGCCCTGAGGCGCTGTGACGAACCTCCAGGGGGCTAAAGGGCTCAGAGTGCAGCCCTGGGGCGCGA

ATG GCC GGG GGC ATC GCT TCC CCG ATG 30 AGC TTC AGC TCT CTC AAG AGG AAG CAA CCC 60 AAG AGC TTC ACC GTG AGG ATC GTC ACC ATG GAC 90 GCC GAG ATG GAG TTC AAT TGC GAG ATG 120 AAG TGG AAA GGG AAG GAC 135
Met Ala Gly Ala Ile Ala Ser Arg Met Ser Phe Ser Ser Leu Lys Arg Lys Gln Pro Lys Thr Phe Thr Val Arg Ile Val Thr Met Asp Ala Glu Met Glu Phe Asn Cys Glu Met Lys Trp Lys Gly Lys Asp

CTC TTT GAT TTG GTG TGC CGG ACT CTG GGG CTC CGA GAA ACC TGG TTC TTT GGA CTG CAG TAC ACA ATC AAG GAC ACA GTG GCC TGG CTC AAA ATG GAC AAG AAG GTA CTG GAT CAT GAT GTT TCA AAG GAA GAA 270
Leu Phe Asp Leu Val Cys Arg Thr Leu Gly Leu Arg Glu Thr Phe Phe Gly Leu Gln Tyr Thr Ile Lys Asp Thr Val Ala Trp Leu Lys Met Asp Lys Val Leu Asp His Asp Val Ser Lys Glu Glu

CCA GTC ACC TTT CAC TTC TTG GCC AAA TTT TAT CCT GAG AAT GCT GAA GAG GAG CTG GTT CAG GAG ATC ACA CAA CAT TTA TTC TTC TTA CAG GTA AAG AAG CAG ATT TTA GAT GAA AAG ATC TAC TGC CCT CCT 405
Pro Val Thr Phe His Phe Leu Ala Lys Phe Tyr Pro Glu Asn Ala Glu Gly Leu Leu Val Gln Glu Ile Thr Gln His Leu Phe Phe Leu Gln Val Lys Lys Gln Ile Leu Asp Glu Lys Ile Tyr Cys Pro Pro

GAG GCT TCT GTG CTC CTG GCT TCT TAC GGC GTC CAG GCC AAG TAT GGT GAC TAC GAC CCC AGT GTT CAC AAG CGG GGA TTT TTG GCC CAA GAG GAA TTG CTT CCA AAA AGG GTA ATA AAT CTG TAT CAG ATG ACT 540
Glu Ala Ser Val Leu Leu Ala Ser Tyr Ala Val Gln Ala Lys Tyr Gly Asp Tyr Asp Pro Ser Val His Lys Arg Gly Phe Leu Ala Glu Glu Leu Leu Pro Lys Arg Val Ile Asn Leu Tyr Gln Met Thr

CGG GAA ATG TGG GAG GAG AGA ATT ACT GCT TGG TAC GCA GAG CAC CGA GGC CGA GCC AGG GAT GAA GCT GAA ATG GAA TAT CTG AAG ATA GCT CAG GAC CTG GAG ATG TAC GGT GTG AAC TAC TTT GCA ATC CGG 675
Pro Glu Met Trp Glu Glu Arg Ile Thr Ala Trp Tyr Ala Glu His Arg Arg Ala Arg Asp Glu Ala Glu Met Glu Tyr Leu Lys Ile Ala Gln Asp Leu Glu Met Tyr Gly Val Asn Tyr Phe Ala Ile Arg

AAT AAA AAG GGC ACA GAG CTG CTG CTT CGA GTG GAT GCC CTG GGG CTT CAC ATT TAT CAC CCT GAG AAC AGA CTG ACC CCC AAG ATC TCC TTC CCG TGG AAT GAA ATC CGA AAC ATC TCC TAC ACT GAC AAG GAG 810
Asn Lys Lys Gly Thr Glu Leu Leu Leu Gly Val Asp Ala Leu Gly Leu His Ile Tyr Asp Pro Glu Asn Arg Leu Thr Pro Lys Ile Ser Phe Pro Trp Asn Glu Ile Arg Asn Ile Ser Tyr Ser Asp Lys Glu

TTT ACT ATT AAA CCA CTG GAT AAG AAA ATT GAT GTC TTC AAG 855 TTT AAC TCC TCA AAG CTT CGT GTT AAT AAG CTG ATT CTC CAG CTA TOT ATC CCG AAC CAT GAT CTA TTT ATG AGG AGA AGG AAA GAT GAT TCT 945
Phe Thr Ile Lys Pro Leu Asp Lys Lys Ile Asp Val Phe Lys Phe Asn Ser Ser Lys Leu Arg Val Asn Lys Leu Ile Leu Gln Leu Cys Ile Gly Asn His Asp Leu Phe Met Arg Arg Arg Lys Ala Asp Ser

TTG GAA GTT CAG CAG ATG AAA GCC CAG CGC AGG GAG AAG GCT AGA AAG CAG ATG GAG CGG CAG CGC CTC GCT CGA GAG AAG CAG ATG AGG GAG GAG GCT GAA CGC ACG AGG GAT CAG TTG GAG AGG AGG CTG 1080
Leu Glu Val Gln Gln Met Lys Ala Gln Ala Arg Glu Glu Lys Ala Arg Lys Glu Arg Leu Ala Arg Leu Ala Arg Glu Lys Gln Met Arg Glu Glu Ala Glu Met Arg Glu Glu Ala Glu Arg GAG GCT

CTG CAG ATG AAA GAA GCA ACA ATG GCG AAC GAA CCA CTG ATG CCG TCT CAG GAG ACA GCT GAC CTG TTG GCT GAA AAG GCC CAG ATC ACC CAG GAG GAG GCA AAA CTT CTG GCC CAG AAG GCC GCA GAG GCT 1215
Leu Gln Met Lys Glu Glu Ala Thr Met Ala Asn Glu Ala Leu Met Arg Ser Glu Glu Thr Ala Asp Leu Leu Ala Glu Lys Ala Gln Ile Thr Glu Glu Glu Ala Lys Leu Leu Ala Gln Lys Ala Ala Glu Ala

GAG CAG GAA ATG CAG CGC ATC AAG GCC ACA GCG ATT CCG CCG GAG GAG CAG AAG CCG CTG ATG GAG CAG AAG GTG CTG GAA GCC GAG GTG CTG GCA CTG AAG ATG GCT GAG GAG TCA GAG AGG AGG GCC AAA GAG 1350
Glu Gln Glu Met Gln Arg Ile Lys Ala Thr Ala Ile Arg Thr Glu Glu Glu Lys Arg Leu Met Glu Gln Lys Val Leu Glu Ala Glu Val Leu Ala Leu Lys Met Ala Glu Glu Ser Glu Arg Arg Ala Lys Glu

GCA GAT CAG CTG AAG CAG GAC CTG CAG GAA GCA CGC GAG GCG GAG CGA AGA GCC AAG CAG AAG CTC CTG GAG ATT GCC ACC AAG CCC ACG TAC CCG CCC ATG AAC CCA ATT CCA GCA CCG TTG CCT CCT GAC ATA 1485
Ala Asp Gln Leu Lys Lys Asp Leu Gln Glu Ala Arg Glu Ala Glu Arg Arg Ala Lys Glu Arg Arg Ala Lys Leu Leu Glu Arg Glu Ala Arg Leu Ala Arg Leu Thr Arg Asp Glu Leu Glu Arg Lys Ile

CCA AOC TTC AAC CTC ATT GGT GAC AGC CTG TCT TTC GAC TTC AAA GAT ACT GAC ATG AAG CCG CTT TCC ATG GAG ATA GAG AAA GAA AAA OTG GAA TAC ATG GAA AAG AGC AAG CAT CTG CAG GAG CAG CTC AAT 1620
Pro Ser Phe Asn Leu Ile Gly Asp Ser Leu Ser Phe Asp Phe Lys Asp Thr Asp Met Lys Arg Leu Ser Met Glu Ile Glu Lys Glu Lys Val Glu Tyr Met Glu Lys Ser Lys His Leu Gln Glu Gln Leu Asn

GAA CTC AAG ACA GAA ATC GAG GCC TTG AAA CTG AAA GAG AGG GAG ACA GCT CTG GAT ATT CTG CAC AAT GAG AAC TCC GAG AGG GGT GGC AGC AGC AAG CAC AAT ACC ATT AAA AAG CTC ACC TTG CAG AGC GCC 1755
Glu Leu Lys Thr Glu Ile Glu Ala Leu Lys Leu Lys Glu Arg Glu Thr Ala Leu Asp Ile Leu His Asn Glu Asn Ser Asp Arg Gly Ser Lys His Asn Thr Ile Lys Lys Leu Ceu Thr Leu Gln Ser Ala

AAG TCC CGA GTG GGC TTC TTT GAA GAG CTC TAG CAGGTGACCCAGCCACCCAGGACCTGCCACTTCTCTGCTACCGGGACCCGGGATGACAGATATCAAGACAGCCATCCATAGGAGCTGGCTGGGGGTTTCCGTGGAGCTCCAGACTTTCCCACT
Lys Ser Arg Val Ala Phe Phe Glu Glu Leu End

FIG. 2 Analysis of deletions in NF2 individuals. *a*, Pulsed-field electrophoresis blots showing rearrangements. From left to right, *Bss*HII digests of a normal control, R2399, R2142 and the affected mother of R2142 hybridized with the probe C16 (upper panel) and the probe C13 (lower panel). C16 detects the expected 130 kb fragment in all individuals; a new 85 kb fragment is seen in individual R2399 and a new 90 kb fragment is seen in individual R2142 and her mother. C13 does not detect the rearranged fragments. *b*, Fluorescent *in situ* hybridization (FISH) of cosmid 101D to the cell line R2142. Top panel, interphase nuclei hybridized with cosmid 101D. Only one chromosome

22 is seen indicating that the cosmid 101D is entirely deleted from the second copy of chromosome 22. Bottom panel, interphase nuclei hybridized with the cosmid 142A. Both copies of chromosome 22 are seen confirming the presence of two intact copies of chromosome 22 at this locus. *c-e*, Nucleotide and predicted amino-acid sequence of N1.1 cDNA, and amino-acid homologies. *c*, Entire nucleotide sequence of the N1.1 cDNA. The predicted start site is at position 1–3 and the in-frame stop codon is at position 1,786–1,788. The predicted amino-acid sequence of 595 residues is shown beneath the respective codons. An in-frame stop codon in the 5' untranslated sequence is 89 ▶

e

Schwa	538	QLNELKTEIEALKLKERETALDI	560
Rat. L1	87	QL+ELK-EIEA+K----ET+LDI	109

METHODS. *a*, High M_r DNA of the lymphoblastoid cell lines was prepared in agarose plugs by standard methods. Restriction digests were done in agarose according to the manufacturer's protocol. The PFGE Pharmacia apparatus at pulse-time settings optimized for separating the desired restriction fragments was used. Southern transfer and hybridizations were as described elsewhere⁴². Lambda ladder M_r markers were purchased from BRL. *b*, Whole cosmid DNA was labelled by nick translation using biotin-14-dATP and hybridized to metaphase chromosome spreads and interphase nuclei of the respective lymphoblastoid cell lines. The signal was detected with Avidin FITC and amplified with Avidin-conjugated antibody⁴¹. *c-e*, The entire N1.1 clone was sequenced in both directions using the dideoxynucleotide method and nested deletion produced using the Henikoff method⁴³. The sequence was further verified by sequencing of genomic clones. The homologies were determined as described in the text.

Isolation of a candidate gene

Southern analysis using the N1.1 cDNA detected seven *EcoRI* fragments. The positions of these *EcoRI* fragments on the cosmid contig show a genomic span of about 100 kb suggesting the presence of many large introns (Fig. 3a).

Tissue specificity and homology

Sequence analysis revealed an open reading frame of 1,785 bases for a predicted protein of 595 amino acids which was named schwannomin, a word derived from schwannoma, the most prevalent tumour seen in NF2 (Fig. 2c). The predicted molecular mass is 66K. The probable start site at position 1-3 is the most likely because it is the first ATG after an in-frame stop codon 89 base pairs (bp) upstream and because it contains the Kozak consensus sequence²⁰. No polyadenylation signal or poly (A) tail was detected. Our 5' untranslated sequence is 144 bp and the 3' untranslated is 135 bp. Based on results obtained with the northern blots, where a 4.5 kb message is detected, we predict that we are missing about 2.3 kb of untranslated sequence of the SCH messenger RNA.

Sequence homology searches using GENBANK and the

BLAST e-mail server indicated that this is a novel sequence²¹. But significant amino-acid homology was found with moesin (48% for entire protein; 62% for the N terminal), ezrin (48% for entire protein; 62% for the N terminal), radixin (47% for entire protein; 62% for the N terminal), protein tyrosine phosphatase 1 (25% for the entire protein; 44% for the N terminal), erythrocyte protein 4.1 (25% for the entire protein; 46% for the N terminal) and rat repetitive element L1 transposon (15 of 22 amino acids in the C terminus; refs 24–27 and W. T. Lankes, H. Furthmayr and M. R. Amieva, sequence submitted to EMBL/Genbank/DDBJ Databanks) (Fig. 2*d,e*). The predicted secondary structure of schwannomin is similar to the structures of moesin, ezrin and radixin: a large N-terminal domain followed by a large α -helix domain and a small C terminus.

Search for point mutations in NF2 patients

To obtain further evidence of SCH alteration in NF2 patients, exons and intron-exons boundaries within the coding sequence of SCH were determined (Fig. 3*a*). Specific exons were amplified by polymerase chain reaction (PCR) and the resulting products were analysed using denaturing gradient gel electrophoresis, an efficient screening method for detecting point mutation²⁸ (Table 2). Survey of a series of 90 unrelated NF2 patients revealed 15 abnormal migration patterns. Direct sequencing of the corresponding bands identified 15 genetic variants. With one exception, which causes a non-conservative

leucine to proline substitution, all these variants can be predicted to lead to the synthesis of a truncated SCH protein (Table 3). Whenever it was possible to investigate several family members in two generations, the SCH mutations were shown to segregate with the disease. On three occasions, the DNA variants were only present in the patient constitutional DNA and not in either of their unaffected parents providing the strongest evidence of a causal relationship between the occurrence of a new mutation and the development of the disease.

Germ-line mutation in patient R2142

To identify the deletion breakpoints of the cell line R2142 we probed *EcoRI* digests with all the *EcoRI* fragments of the SCH region. Variant *EcoRI* bands were only identified with the two *EcoRI* fragments shown in Fig. 3*a*, providing evidence that the borders of the deletion were lying within these two fragments and predicting a 36 kb deletion compatible with the conclusion derived from the previous PFGE and FISH studies. This deletion, entirely contained within the SCH gene, is predicted to remove the second, third and fourth coding exon. Reverse-transcriptase PCR of RNA isolated from the lymphoblastoid cell line R2142, using primers internal to the cDNA yielded in addition to the expected 764 bp fragment, a smaller 431 bp fragment. Sequencing revealed that the latter fragment originated from a SCH mRNA in which the first and the fifth coding exon were adjacent thus confirming the initial prediction. The

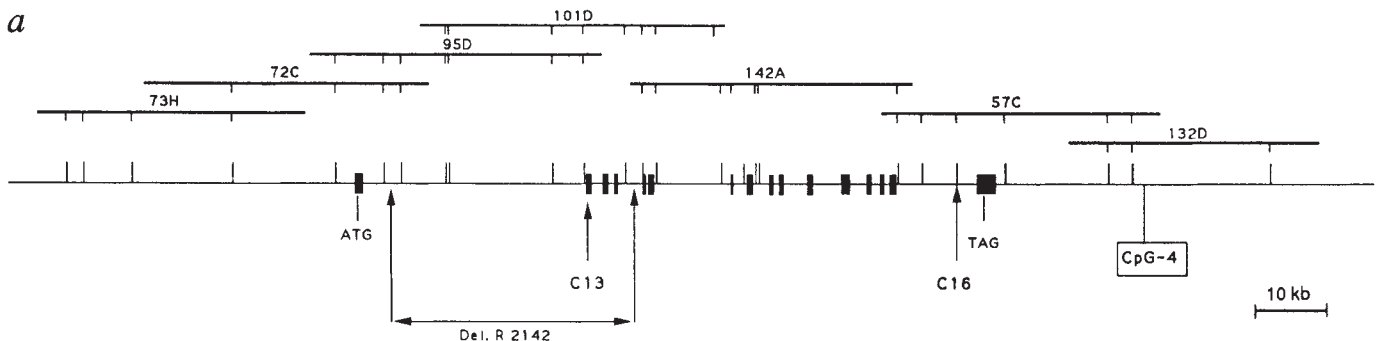


FIG. 3 *EcoRI* restriction map of the SCH gene, position of coding exons and characterization of the aberrant transcript in patient R2142. *a*, Vertical bars and black boxes indicate the position of *EcoRI* sites and of the coding exons, respectively. The first and last coding exons are indicated by ATG and TAG, respectively. The position of each exon within its *EcoRI* fragment is only indicative. The positions of the C13 and C16 probes are also provided. The position of the R2142 deletion is shown. The positions of the two *EcoRI* fragments detecting rearranged bands in patient R2142 DNA are marked by arrows. *b*, The sequence of the reverse transcriptase PCR (RT-PCR) products from R2142 RNA. Comparison to the N1.1 sequence shows an in-frame deletion of 333 bp of the expressed SCH gene in R2142. **METHODS.** The *EcoRI* restriction map of the 7 cosmids was generated using standard procedures⁴². *Sau3A*-digested cosmid DNAs were also subcloned in M13 mp18 and the resulting mini-libraries were screened with the N1.1 cDNA. The positive clones were directly sequenced using dideoxy chain termination methods and an Applied Biosystems automatic sequencer. Missing exons were searched among restriction fragments generated by appropriate digestion of the cosmid DNAs using N1.1 cDNA subfragments as probes. The sequence of the intron-exon junctions are deposited in the EMBL data base (accession number X7265 HSSCH01 to X72670 HSSCH16). Total RNA was prepared from the R2142 lymphoblastoid cell line by standard methods. RT-PCR was done using the primers TCTCAAGAGGAAG-CAACCCA and CGCTGTACGAGATGTTTCGG (double underlined in Fig. 2*c*), the Perkin-Elmer GeneAmp kit and the manufacturer's protocol. The annealing temperature was 55 °C for 1 min, elongation at 72 °C for 1.5 min and denaturation at 94 °C for 1 min for 30 cycles. The expected 764 bp RT-PCR product plus a new 431 bp product were visualized on agarose gels. The 431 bp fragment was purified and sequenced after asymmetric PCR amplification.

b R2142 deletion

N1.1 TCTCAAGAGGAAGCAACCCAAGACGTTACCGTGAGGATCGTCACCATGGAAGCCGAGAT
R2142 TCTCAAGAGGAAGCAACCCAAGACGTTACCGTGAGGATCGTCACCATGGAAGCCGAGAT

N1.1 GGAGTTCAATTGCGAGATGAAGTGGAAAGGGAAGGACCTCTTTGATTGGTGTCCGGAGT
R2142 GGAGTTCAATTGCGAG

N1.1 CTGGGGCTCCGAGAAACCTGGTCTTTGGACTGCAGTACACAATCAAGGACACAGTGGCCT
R2142

N1.1 GGCTCAAAATGGACAAGAAGCTACTGGATCATGATGTTTCAAAGGAAGAACCAAGTCAAC
R2142

N1.1 TTTCACTCTTGGCCAAATTTATCCTGAGAATGCTGAAGAGGAGCTGGTTCAGGAGATCA
R2142

N1.1 CACAACATTTATTCTTCTACAGGTAAGAAGCAGATTTATAGTGAAGATCTACTGC
R2142

N1.1 CCTCTGAGGCTTCTGTCTCTGGCTCTTACGCCGTCCAGGCCAAGTATGGTGACTACGAC
R2142 TATGGTGACTACGAC

N1.1 CCCAGTGTTCAAGCGGGGATTTTGGCCCAAGAGGAATTGCTTCAAAAAGGGTAATA
R2142 CCCAGTGTTCAAGCGGGGATTTTGGCCCAAGAGGAATTGCTTCAAAAAGGGTAATA

N1.1 AATCTGTATCAGATGACTCCGGAATGTGGGAGGAGAGAATTACTGCTTGGTACGCAGAG
R2142 AATCTGTATCAGATGACTCCGGAATGTGGGAGGAGAGAATTACTGCTTGGTACGCAGAG

N1.1 CACCGAGGCCGAGCCAGGATGAAGCTGAAATGGAATATCTGAAGATAGCTCAGGACCTG
R2142 CACCGAGGCCGAGCCAGGATGAAGCTGAAATGGAATATCTGAAGATAGCTCAGGACCTG

N1.1 GAGATGTACGGTGTGAACACTTTGCAATCCGGAATAAAAAGGGCAGAGCTGCTGCTT
R2142 GAGATGTACGGTGTGAACACTTTGCAATCCGGAATAAAAAGGGCAGAGCTGCTGCTT

N1.1 GGAGTGGATGCCCTGGGCTTACATTTATGACCTGAGAACAGACTGACCCCAAGATCT
R2142 GGAGTGGATGCCCTGGGCTTACATTTATGACCTGAGAACAGACTGACCCCAAGATCT

N1.1 CCTTCCCGTGAATGAAATCCGAAACATCTCGTACAGTG
R2142 CCTTCCCGTGAATGAAATCCGAAACATCTCGTACAGTG

reading frame is preserved in this aberrant mRNA and its predicted translation product is a schwannomin protein in which amino acids 39 to 149 are deleted (Fig. 3b). Because patient R2142 is affected, it is expected that this abnormal SCH protein is non-functional.

Search for SCH mutations in NF2-related tumours

DNA from 30 schwannomas and 30 meningiomas were also investigated for mutations in the same 6 coding exons. Somatic SCH mutations were observed in two schwannomas (tumours IC-J and IC-F) and two meningiomas (tumours IC-AB and IC-PO) which developed in patients not affected by NF2. In addition, in two schwannomas (tumours 28A and IC-BE) from NF2 patients, the SCH germ-line mutations were detected (Table 3). Loss of heterozygosity for chromosome 22 alleles in tumours IC-J, IC-F and IC-PO indicate that both copies of SCH are inactivated in three tumours. The absence of a normal SCH allele in tumour 28A and detection of the germ-line SCH mutation in the remaining chromosome leads to the same conclusion for a fourth tumour. Thus functional inactivation of SCH by a two-hit mechanism may operate in the tumorigenic process of both schwannomas and meningiomas.

Discussion

The homology of schwannomin to erythrocyte protein 4.1 and ezrin/moesin/talin family of genes suggests that schwannomin sublocalizes to the cell membrane and acts as a membrane-organizing protein. From these homologies we predict that the N-terminal domain binds to integrins and the α -helix domain binds to components of the cytoskeleton^{23,25,29}. The protein may normally act like protein 4.1 which links transmembrane glycoproteins to the spectrin-actin complex of the cytoskeleton or, like talin, which interacts with vinculin and integrins, thereby regulating organization of cell shape and perhaps cytoplasmic extensions. For example, schwannomin might normally lead to stable cell-cell and cell-matrix interactions. Its absence may

lead to cell migration, changes in cell shape or loss of contact inhibition. Ezrin and moesin are found preferentially in retraction fibres, blebs, microspikes, filopodia and lamellipodia, structures involved in cell exploration, attachment, movement, and events in epithelial-mesenchymal transformations in development.

Clearly, the elucidation of the role of schwannomin in the development of the NF2 phenotype awaits further analysis of its biochemistry and cellular biology. However, identification in NF2 patients of SCH mutations which almost always lead to a truncated protein, or of deletion involving a part of or a complete copy of the SCH gene, clearly indicates that loss of one functional copy of SCH causes the disease. The identification of potentially inactivating somatic SCH mutations in tumours known to occur in NF2 patients further substantiates the hypothesis that SCH has tumour-suppressor activity. Deletion studies have implicated a tumour-suppressor gene on chromosome 22 in the genesis of a group of tumours not seen with a high frequency in NF2: gliomas³⁰, pheochromocytomas³¹, colon carcinoma³² and breast cancer³³. Genotypic analysis of such tumours known to lose chromosome 22 frequently will determine more precisely the tumour types in which functional inactivation of SCH may occur and thus document further the range of its potential tumour-suppressor activity.

Finally the demonstration that SCH mutations are associated with the NF2 phenotype has immediate implications. First, members of NF2 families can now be directly tested for mutations in this gene. Those individuals that have not inherited the gene will be spared repeated medical evaluation. Second, the phenotypic manifestations of NF2, which vary widely among patients and families³⁴, can now be analysed on the basis of fully characterized SCH mutation.

Note added in proof: Since submission of this article a candidate gene for NF2 has been reported by Trofatter *et al.* (*Cell* **72**, 791-800, 1993). Its cDNA and deduced translation product, termed Merlin, are almost identical to the SCH cDNA and to the schwannomin protein. □

Received 2 March; accepted 15 April 1993.

1. Tos, M. & Thomsen, J. *J. Lar. Otol.* **98**, 685-692 (1984).
2. Martuza, R. L. & Eldridge, R. N. *Engl. J. Med.* **318**, 684-688 (1988).
3. Barker, D. *et al. Science* **236**, 1100-1102 (1987).
4. Seizinger, B. R. *et al. Cell* **49**, 589-594 (1987).
5. Rouleau, G. A. *et al. Nature* **329**, 246-248 (1987).
6. Narod, S. A. *et al. Am. J. hum. Genet.* **51**, 486-496 (1992).
7. Seizinger, B. R., Martuza, R. L. & Gusella, J. F. *Nature* **332**, 644-647 (1986).
8. Couturier, J. *et al. Cancer Genet. Cytogenet.* **45**, 55-62 (1990).
9. Seizinger, B. R., de la Monte, S., Atkins, L., Gusella, J. F. & Martuza, R. L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5419-5423 (1987).
10. Dumanski, J. P., Carlsson, E., Collins, V. P. & Nordenskjöld, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9275-9279 (1987).
11. Dumanski, J. P., Rouleau, G. A., Nordenskjöld, M. & Collins, V. P. *Cancer Res.* **50**, 5863-5867 (1990).
12. Rutledge, M. H. *et al. Neurology* (in the press).
13. Delattre, O. *et al. Genomics* **9**, 721-727 (1991).
14. Dumanski, J. P. *et al. Hum. Genet.* **84**, 219-222 (1990).
15. Zucman, J. *et al. Genes Chrom. Cancer* **5**, 271-277 (1992).
16. Lindsay, S. & Bird, A. P. *Nature* **327**, 336-338 (1987).
17. Delattre, O. *et al. Nature* **359**, 162-165 (1992).
18. Mattei, M. G. *et al. Hum. Genet.* **80**, 293-295 (1988).
19. Lees, J., Shneiderman, P. S., Skuntz, S. F., Carden, M. J. & Lazzarini, R. A. *EMBO J.* **7**, 1947-1955 (1988).
20. Kozac, M. *Nucleic Acids Res.* **16**, 8125-8130 (1988).
21. Aitshul, S. F. *et al. J. molec. Biol.* **215**, 403-410 (1990).
22. Lankes, W. T. & Furthmayr, H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8297-8301 (1991).
23. Gould, K. L., Bretscher, A., Esch, F. S. & Hunter, T. *EMBO J.* **8**, 4133-4142 (1989).
24. Turunen, O. *et al. J. biol. Chem.* **264**, 16727-16732 (1989).
25. Yang, Q. & Tonks, N. K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5949-5953 (1991).
26. Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9512-9516 (1986).
27. Kahre, O., Ilves, H. & Speck, M. *Molec. cell. Biol.* **12**, 4242-4248 (1992).
28. Myer, R. M., Fisher, S. G., Lerman, L. S. & Maniatis, T. *Nucleic Acids Res.* **13**, 3131-3145 (1985).
29. Rees, D. J. G., Ades, S. E., Singer, S. J. & Hynes, R. O. *Nature* **347**, 685-689 (1990).

30. Jenkins, R. B. *et al. Cancer Genet. Cytogenet.* **39**, 253-279 (1989).
31. Tanaka, N. *et al. Genes Chrom. Cancer* **5**, 399-403 (1992).
32. Okamoto, M. *et al. Nature* **331**, 273-277 (1988).
33. Chen, L. C., Kurisu, W., Ngo, J., Moore, D. & Smith, H. S. *Cancer Res.* **51**, 300 (1991).
34. Evans, D. G. R. *et al. J. med. Genet.* **29**, 841-846 (1992).
35. Lerman, L. S. & Silverstein, K. *Meth. Enzym.* **155**, 482-501 (1987).
36. NIH Consensus Dev Conf Consensus Statement **9(4)**, Dec 11-13 (1991).
37. van Leewen, C. *et al. Nucleic Acids Res.* **19**, 5805 (1991).
38. Spirio, L., Joslyn, G., Nelson, L., Leppert, M. & White, R. *Nucleic. Acids Res.* **19**, 6348 (1991).
39. Fontaine, B. *et al. Genomics* **10**, 280-283 (1992).
40. Bijlsma, E. *et al. Genes Chrom. Cancer* **5**, 201-205 (1992).
41. Desmaze, C., Zucman, J., Delattre, O., Thomas, G. & Aurias, A. *Genes Chrom. Cancer* **5**, 30-34 (1992).
42. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
43. Henikoff, S. *GENE* **28**, 357 (1984).
44. Olschwang, S., Laurent-Puig, P., Groden, J., White, R. & Thomas G. *Am. J. hum. Genet.* **52**, 273-279 (1993).

ACKNOWLEDGEMENTS. We thank M. Giovannini and L. Selleri for their contribution to YAC isolation and characterization, and T. Melot, M. Peter, A. Krizus, L. Lebus, M. Rousseau and J. Sarrazin for technical assistance; K. Morgan and M. Fujiwara of the NCE medical genetics computer unit for their assistance with the database searches; F. Resche, J. M. Sterkers, J. Philippon, P. Vialat and J. F. Gusella for providing cell lines or samples; and the NF2 patients and their families for their cooperation. This work was supported by grants from the European Community Commissions, the Ministère de la Recherche et de la Technologie, the Ligue Nationale contre le Cancer, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale, the Association Française contre les Myopathies, the Council for Coordinating Cancer Research, the Association Neurofibromatose France, Medical Research Council of Canada, the National Neurofibromatosis Foundation Inc. and the Fonds de Recherche en Santé du Québec, the A. & M. Axson Johnson's Fond, the Swedish Medical Research Council, S. and L. Walker Foundation and the National Institutes of Health. K.H.X., C.D. and E.B. are recipient of fellowships from Comité Départemental de l'Yonne de la Ligue Contre le Cancer and from Association pour la Recherche sur le Cancer and from the Dutch Praeventiefonds respectively.

Observations of high-velocity, weakly shocked ejecta from experimental impacts

Andrew J. Gratz, William J. Nellis & Neal A. Hinsey

Institute of Geophysics and Planetary Physics, and H Division, Lawrence Livermore National Laboratory, University of California, Livermore, California 94550, USA

A SMALL proportion of meteorites found on Earth are thought to come from planet-sized bodies^{1,2}. The 'lunar meteorites' are now well established as having come from the Moon³⁻⁶ on the basis of direct comparison with lunar samples. The SNC meteorites (shergottites, nakhlites and chassignites) — seven achondrite meteorites distinguished by extremely young formation ages (<1.3 Gyr), high volatile contents, distinctive oxygen isotopic ratios and rare earth compositions — are igneous rocks, believed² to have formed on a planet, probably Mars. But it is hard to reconcile the weakly shocked nature of many lunar and SNC meteorites with the strong shock metamorphism known to accompany impacts of the size required to eject material from a planet-sized body. Computer modelling⁷⁻¹⁰ of impacts has yet to resolve this issue, although it has been proposed^{9,10} that surface rarefaction near an impact can produce high-velocity, weakly shocked ejecta. Here we present the results of a cratering experiment which separates and captures the ejecta from different regions around the impact site. We recover high-velocity, weakly shocked material as predicted^{9,10}, lending additional support both to our understanding of cratering mechanics and to a planetary or lunar origin for the SNC and lunar meteorites.

Shock pressure in lunar meteorites is generally reported as < 20 GPa (200 kbar), whereas that in SNCs is generally higher: 30 GPa for shergottites, 0 GPa for nakhlites, and 0–175 GPa for chassignites (see ref. 2). Local stresses can be higher or lower. The variations in different groups may be linked to details of launch sites and ejection processes. By comparison, strongly shocked impactites are completely amorphized or melted (45 to 100 GPa).

If these meteorites do originate by ejection from other planets, they probably come from near the surface close to the crater, which is ejected at high velocity and is weakly shocked. In a meteorite impact, shock pressure falls off rapidly with distance from the impact site (Fig. 1), generating strongly shocked, high-speed ejecta close to the impact and weakly shocked ejecta below and outwards. Most of this weakly shocked material ends up as crater fill or as local ejecta.

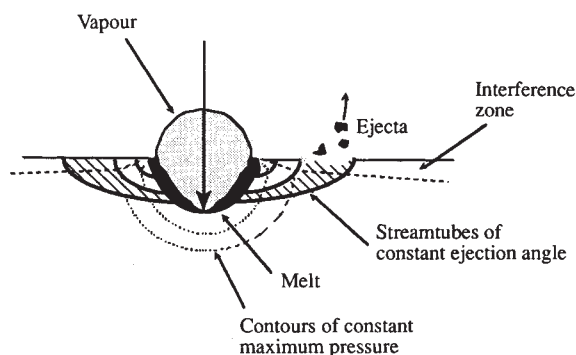


FIG. 1 Schematic of a meteorite impact, showing the near-surface interference along with pressure contours. After ref. 14.

However, there is also an 'interference zone' near the surface in which the surface rarefaction releases the target before maximum shock pressure is reached⁹. This region, of thickness roughly 20% of the projectile radius close to the impact and thicker farther away, is ejected at the highest velocities of any ejecta but experiences shock pressures $\leq 1\%$ of the impact pressure. Thus, we expect most of the longest-range ejecta to be relatively unshocked, even though the interference zone accounts for 10% of the total ejecta. Target materials from intermediate depths experience a wide range of pressures and may become either ejecta or crater fill.

Although many cratering experiments had been done before the SNC controversy arose, they did not fully test the ejection hypothesis in a compact target. A study of a very large crater in granite¹¹ examined the crater ejecta, but no distinction was made between interference-zone ejecta and other ejecta. High-speed photography of spall plates (material ejected from the surface itself) around an experimental crater¹² obtained data on plate velocity during impact, but could not resolve fragments much smaller than the impactor itself. A third study¹³ examined ejecta from impacts in unconsolidated sand. The loading and ejection history in sand will differ considerably from that of planetary surfaces which, except for a thin veneer, are strong and cohesive.

A practical difficulty in examining ejecta from craters in or out of the laboratory is that ejecta become mixed during crater excavation. Numerical studies show that shock pressure contours are concentric about the impact site, whereas material reaching any given point around a crater comes from an ejection streamtube (region of constant ejection angle) extending more-or-less radially outward from the impact¹⁴ (Fig. 1). Around large craters, the ejecta record may be further complicated by debris flows and atmospheric interactions. Strongly and weakly deformed material can thus be co-deposited¹⁵.

Our experiment selectively separates and captures ejecta from the interference zone. A 3-mm-thick, 19-mm-diameter, flat aluminium flyer plate on a lexan sabot 20 mm in diameter is accelerated to 4.0 km s⁻¹ using a 6.5-m-long two-stage light-gas gun. The projectile travels down an entrance tube cut in a foam cylinder (the recovery fixture), and strikes a stationary disk of Westerly granite, 150 mm in diameter and 100 mm thick (Fig. 2). The impact pressure is 43 GPa (430 kbar), calculated from the measured projectile velocity, the equations of state of Al and granite, and the standard method of matching shock-impedance¹⁶. The impact causes material in the interference zone about the disk to be ejected backward and into the foam cylinder; material under the projectile is either trapped or scattered forward. In a meteorite impact, this strongly shocked

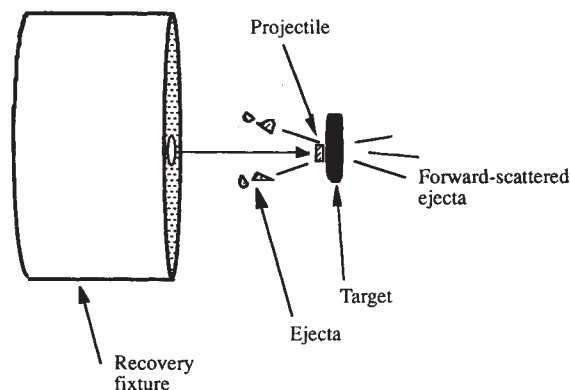


FIG. 2 Design of experiment showing projectile travelling down a hollow tube on the axis of a CH foam (density 0.03 g cm⁻³) recovery fixture and impacting sample. Captured ejecta are emitted backwards in the direction of the incident projectile and cut a thin-walled conical section in the foam.

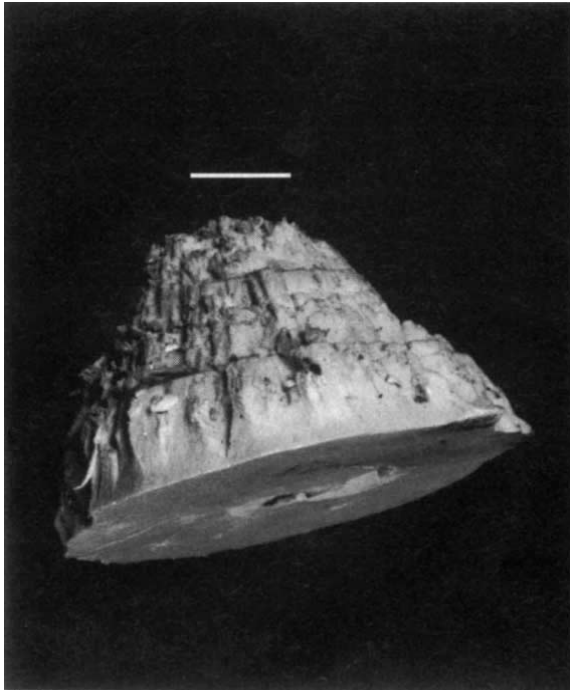
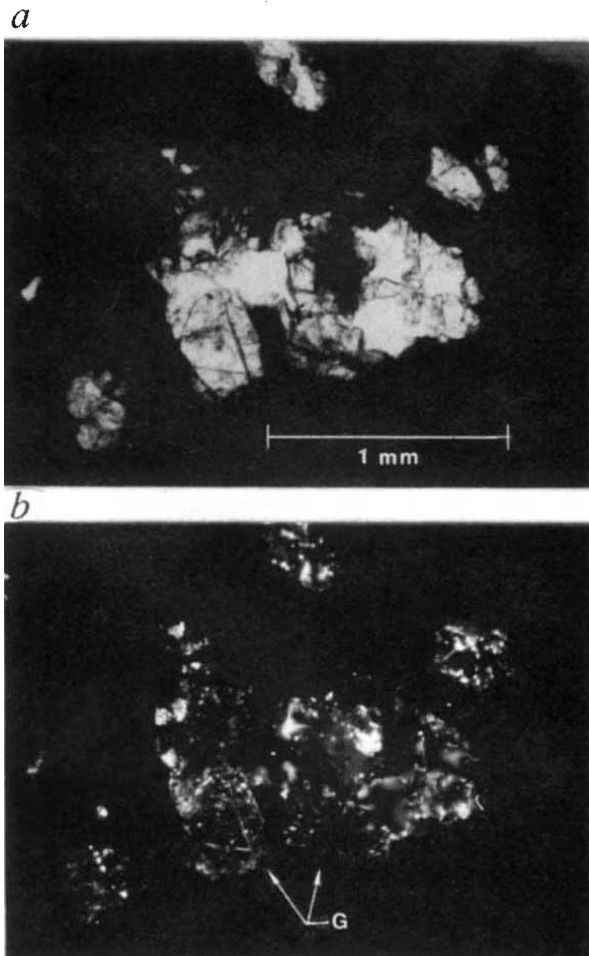


FIG. 3 Cone (half-angle 40°) cut in the foam recovery fixture by high-speed ejecta. The scale bar is 50 mm.



material would rebound and eject upward at high speed, mixing with material from the interference zone. In our experiments, only a small amount ($\ll 0.1$ g) of this material escapes from beneath the projectile to be propelled back along the projectile path and lands on the gun muzzle. An optical microscope is used to examine material embedded in the flyer plate, caught in the foam or caught on the gun muzzle.

We selectively capture the interference zone ejecta in a recovery fixture which is 4 cm away from the sample and made of low-density CH foam to minimize post-ejection damage. Even for ejecta travelling at the impact velocity of 4 km s^{-1} , the shock stress caused by impact with the foam would generate stresses well below 10 GPa, the nominal threshold for inducing planar deformation features (PDFs) and most other shock indicators in quartz and feldspars^{17,18}. Very slow ($< 100 \text{ m s}^{-1}$) particles cannot penetrate the foam at all¹⁹.

The recovered ejecta takes the form of particles mostly 0.1 to 4 mm in diameter, travelling along the thin outer surface of a cone, and indeed cuts out an ejecta cone in the foam with a half-angle of $\sim 40^\circ$ for an ejecta elevation angle of $\sim 50^\circ$ (Fig. 3). The foam penetration depths indicate¹⁹ ejection velocities of about 1 km s^{-1} . Velocities, ejection angles and particle sizes are in general agreement with numerical predictions⁹ for interference-zone material, suggesting that our method of isolating interference-zone ejecta is successful.

Rock embedded in the Al flyer plate is obviously shocked, containing abundant diaplectic glass, PDFs in quartz, and extensive fracturing and granulation of quartz and feldspars (Fig. 4a, b). On the basis of shock piezometric indicators²⁰, a range of shock pressures up to ~ 40 GPa is generated, in agreement with the shock impedance calculation. Ejecta deposited in the backscattered direction on the gun muzzle is shocked (Fig. 4c) and was reduced to grains mostly $< 50 \mu\text{m}$ in diameter.

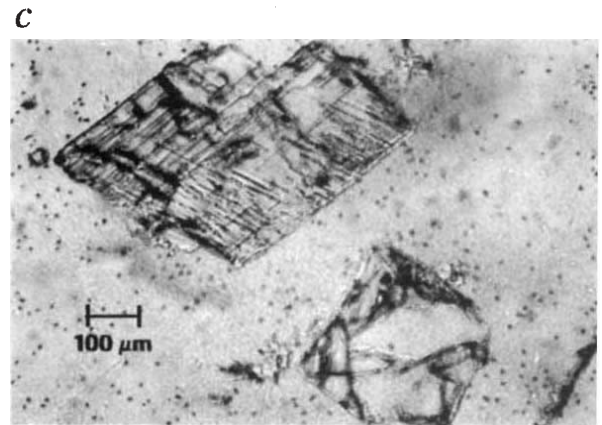


FIG. 4 Shock features seen in recovered material. Diaplectic glass representing > 40 GPa pressures is embedded in the aluminium flyer plate: glass areas (G) are clear in polarized light (a), dark in crossed polars (b). Planar fractures (c) are present in feldspar from the ejecta cone and indicate shock pressures below 10 GPa.

In contrast, rock from the thin-walled ejecta cone shows little or no sign of shock features. There are no amorphized regions or PDFs, and only occasional planar fractures; the latter are only in feldspar, not in quartz. Grains are large, ranging from tens of micrometres to several millimetres. Fracturing was common, much of it intergranular. By comparison with moderate²¹ and low-pressure²² shock recovery experiments, we estimate that the peak pressure experienced by the ejecta is <5 GPa for most of the material, 5–10 GPa for <10% of it. Thus, most interference-zone ejecta experience peak pressures of only a few per cent of the impact pressure.

Because gravity has a negligible effect on high-speed ejecta at short range, our experiments can be scaled directly to larger events by using numerical codes. These calculations show that crater features scale linearly with projectile size, so larger ejecta cones would be similar to those in our experiments. In addition, modelling⁹ predicts that fragments as large as 4–40 m in diameter could be ejected from (and escape) the Moon; 1-m blocks could escape from Mars. Scaling in projectile velocity is more difficult. However, for representative hypervelocity impacts of 15–30 km s⁻¹ (impact pressures >100 GPa) much of the interference-zone ejecta should still be unmelted and only weakly shocked: 1% of 150 GPa is 1.5 GPa, at the lower limit for detectable shock metamorphism.

Our results support theories of lunar and martian origins for some meteorites. Their source is the interference zone from which they escape at high velocity but relatively unshocked. At the higher impact velocities representative of meteorite impacts, shock levels will be higher on average but still a small fraction of the impact pressure. In contrast, ejecta originating from beneath the point of impact of a meteorite are highly shocked, as

expected. The same mechanisms are expected to apply for all planets, moons and asteroids in the Solar System. Our results indicate that meteorites reaching the Earth from the planets and moons will be dominantly composed of high-speed, weakly shocked ejecta from the interference zone close to the impact site. □

Received 18 January; accepted 20 April 1993.

1. Wasson, J. T. & Wetherill, G. W. in *Asteroids* (ed. Gehrels, T.) 926–974 (Pergamon, Oxford, 1979).
2. Wood, C. A. & Ashwal, L. D. *Proc. Lunarplanet. Sci. Conf.* **12B**, 1359–1375 (1981).
3. Warren, P. H., Taylor, G. J. & Keil, K. *Geophys. Res. Lett.* **10**, 779–782 (1983).
4. Mayeda, T. K., Clayton, R. N. & Molini-Velsko, C. A. *Geophys. Res. Lett.* **10**, 799–800 (1983).
5. Bogard, D. D. & Johnson, P. *Geophys. Res. Lett.* **10**, 801–803 (1983).
6. Paime, H., Sperel, B., Weckwerth, G. & Wänke, H. *Geophys. Res. Lett.* **10**, 817–820 (1983).
7. O'Keefe, J. D. & Ahrens, T. J. *Icarus* **62**, 328–338 (1985).
8. Vickery, A. M. & Melosh, H. J. *Icarus* **56**, 299–318 (1983).
9. Melosh, H. J. *Icarus* **59**, 234–260 (1984).
10. Melosh, H. J. *Geology* **13**, 144–148 (1985).
11. Hörz, F. *Contr. Mineral. Petrol.* **21**, 365–377 (1969).
12. Polanskey, C. A. & Ahrens, T. J. *Icarus* **87**, 140–155 (1990).
13. Stöffler, D., Gault, D. E., Wedekind, J. & Polkowski, G. J. *geophys. Res.* **80**, 4062–4077 (1975).
14. Melosh, H. J. *Impact Cratering: A Geologic Process* 1–245 (Oxford Univ. Press, 1989).
15. Hörz, F., Ostertag, R. & Rainey, D. A. *Rev. Geophys. Space Phys.* **12**, 309–327 (1983).
16. Altschuler, L. V., Krupnikov, K. K., Ledenev, B. N., Zhuchikhin, V. I. & Brazhnik, M. I. *Sov. Phys. JETP* **7**, 606–614 (1958).
17. Hörz, F. in *Shock Metamorphism of Natural Materials* (eds French, B. M. & Short, N. M.) 243–254 (Mono, Baltimore, MD, 1968).
18. Müller, W. F. & Defourneaux, M. W. *Geophys.* **34**, 483–504 (1968).
19. Hörz, F., Fujiwara, A. & Fujii, N. *Jap. J. appl. Phys.* **129**, 2543–2549 (1990).
20. Stöffler, D. *Fortschr. Miner.* **49**, 50–113 (1972).
21. Gratz, A. J., et al. *Phys. Chem. Minerals* **19**, 267–288 (1992).
22. Gratz, A. J., Nellis, W. J. & Hinsey, N. A. *Geophys. Res. Lett.* **19**, 1391–1394 (1992).

ACKNOWLEDGEMENTS. We thank P. Warren for helpful discussions and advice and F. Ryerson for providing the granite. This work was done under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory.

Imaging standing waves in a two-dimensional electron gas

M. F. Crommie, C. P. Lutz & D. M. Eigler

IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, California 95120, USA

ELECTRONS occupying surface states on the close-packed surfaces of noble metals form a two-dimensional nearly free electron gas^{1–3}. These states can be probed using the scanning tunnelling microscope (STM), providing a unique opportunity to study the local properties of electrons in low-dimensional systems⁴. Here we report the direct observation of standing-wave patterns in the local density of states of the Cu(111) surface using the STM at low temperature. These spatial oscillations are quantum-mechanical interference patterns caused by scattering of the two-dimensional electron gas off step edges and point defects. Analysis of the spatial oscillations gives an independent measure of the surface state dispersion, as well as insight into the interaction between surface-state electrons and scattering sites on the surface.

The experiments were done using a STM contained in ultrahigh vacuum and cooled to 4 K (ref. 5). The single-crystal Cu sample was prepared by repeated cycles of Ar ion sputtering and annealing, before being cooled. A polycrystalline tungsten wire was used for the tip of the STM. The convention used here is that the bias across the tunnel junction (V) is the voltage of the sample measured with respect to the tip.

Figure 1 shows a constant current $500 \text{ \AA} \times 500 \text{ \AA}$ STM image of the Cu(111) surface, acquired with a bias of $V=0.1 \text{ V}$. At this bias the STM image roughly corresponds to a surface of constant local density of states (LDOS) at the Fermi energy (E_F)⁶. The most striking aspect of this image is the existence of static spatial oscillations in the Cu(111) LDOS. The oscillations decay away

from step edges and point defects, and have a characteristic period of $\sim 15 \text{ \AA}$. The amplitude of the corrugation is $\sim 0.02 \text{ \AA}$ and is greater near the top of step edges than near the bottom. Other images were recorded (at lower junction resistance) that simultaneously displayed both the LDOS oscillations and the underlying Cu lattice. The origin of the point defects visible in Fig. 1 is unknown, although Auger analysis of the sample during preparation indicates that they are probably due to trace amounts of sulphur or antimony contaminants.

The differential conductance (dI/dV) of a STM tunnel junction is closely related to the LDOS of the surface under the tip⁷. Spatial variations in the LDOS of a surface at the energy $E=E_F+|e|V$ can be approximately mapped out by measuring dI/dV over the surface at bias voltage V (ref. 8). (One source of error in this interpretation comes from tip-height variations during dI/dV measurement. This error can be ignored when the fractional change in exponential barrier penetration factor⁷ is negligible compared with the fractional change in surface LDOS during a dI/dV measurement.) Figure 2a shows the measured variation in dI/dV as a function of distance from a monatomic step. Measurements were taken while scanning the tip laterally across the surface and adjusting the tip height to maintain constant d.c. current. Bias voltages were chosen to probe states both above and below E_F . Each linescan shown in Fig. 2a was measured over the same small patch of surface and is the average of many scans. The linescans show a local minimum in dI/dV that occurs roughly halfway up the step ($x=5 \text{ \AA}$), although the uncertainty in this position is $\pm 3 \text{ \AA}$.

The dI/dV linescans in Fig. 2a reveal that the wavelength of the oscillation in surface LDOS changes as a function of energy. The wavelength smoothly increases as energy is lowered with respect to E_F . At $\sim 0.45 \text{ eV}$ below E_F a transition occurs where the LDOS sharply decreases in magnitude. Some residual spatial oscillations in the dI/dV scans are still observed below this energy. These do not change wavelength with energy, and are probably caused by tip-height variations.

Similar behaviour is seen in the circular LDOS oscillations emanating from point defects on the surface. Figure 2b shows the spatial variation in tip height obtained from a constant current linescan across one such defect. For the low bias voltage used here (20 mV) and small tip-height variations (~ 0.02 Å), changes in tip height across the defect can be interpreted as proportional to variations in the surface LDOS at E_F . Spatial oscillations with a period of ~ 15 Å are clearly evident.

Complementary information regarding the LDOS of the surface is obtained by measuring a full dI/dV spectrum at one point on the surface. Spectra were recorded by holding the STM tip stationary over various spots on the Cu surface and recording

dI/dV while ramping the bias from 1.0 to -1.0 V at each spot. Figure 3 shows normalized dI/dV spectra measured while centring the STM tip first over a clean section of terrace (where 'clean' means that the tip is at least 40 Å from any step or point defect) and then over the centre of a monatomic step edge. The tip remained unchanged between these measurements. The most dramatic feature in the clean terrace spectrum is the sharp drop, or cusp, in dI/dV at $V \approx -0.45$ V. This corresponds to a sudden decrease in the surface LDOS at energies more than 0.45 eV below E_F . The clean terrace dI/dV also displays an oscillatory component, as one might expect given the energy-dependent wavelength of the spatial LDOS oscillations. The spectrum

FIG. 1 Constant-current $500 \text{ Å} \times 500 \text{ Å}$ image of the Cu(111) surface ($V=0.1$ V, $I=1.0$ nA). Three monatomic steps and about 50 point defects are visible. Spatial oscillations with a periodicity of ~ 15 Å are clearly evident. The vertical scale has been greatly exaggerated to display the spatial oscillations more clearly.

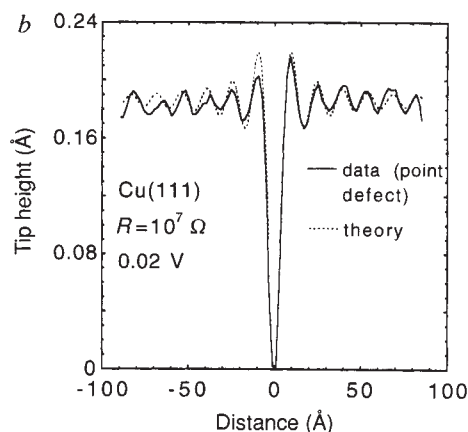
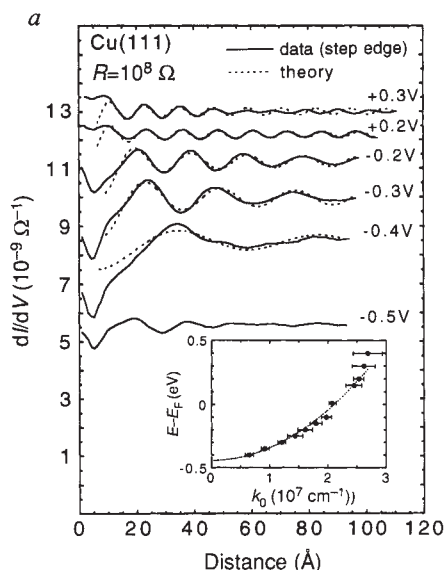
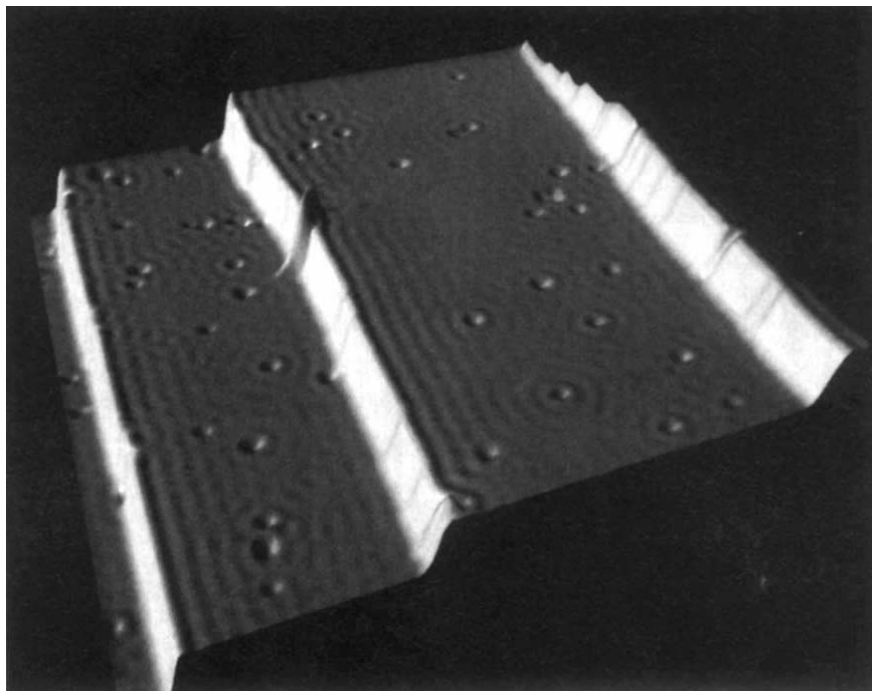


FIG. 2 *a*, Solid lines: spatial dependence of dI/dV , measured as a function of distance (along upper terrace) from step edge at different bias voltages. Zero distance corresponds to the lower edge of the step.

The d.c. junction resistance was kept at $10^8 \Omega$ for each linescan (a 10-mV, 205 r.m.s.-Hz sine wave was used for a.c. measurement of dI/dV). Dashed lines: theoretical fits of equation (3) to data. Curves have been shifted vertically for viewing. Inset: experimental surface-state dispersion, obtained from fits of equation (3) to dI/dV linescan data. Dashed line is the parabolic fit used to extract the surface state effective mass and band edge. *b*, Solid line: height of STM tip during linescan across point defect ($V=0.02$ V, $I=2.0$ nA). Dashed line: fit of equation (5) to data.

measured over the step edge differs in that there is no sharp decrease at $V \approx -0.45$ V. dI/dV spectra taken over point defects similarly showed a smaller cusp feature than spectra obtained over clean sections of terrace. Similar results were obtained even when the electronic properties of the tip were changed by touching the tip to the surface (data not shown).

The dI/dV spectra shown in Fig. 3 are similar to dI/dV spectra measured on Au(111) at room temperature by Everson *et al.*⁸. They observed a similar cusp feature in spectra taken over clean Au terraces, and a suppression of that feature in spectra taken over monatomic Au steps. Their results were explained in terms of two-dimensional surface states. Surface states arise on the (111) face of the noble metals as a result of the gap that exists along the Γ -L line in their bulk band structure^{9,10}. Electrons occupying such states are bound in the direction perpendicular to the surface, but have free-electron-like characteristics parallel to the surface. Everson *et al.*⁸ explained the sharp cusp in their clean-terrace dI/dV spectra as marking the bottom of such a surface-state band on Au(111), and the suppression of this feature at steps as a reduction in surface-state LDOS due to a repulsive barrier at step edges. Davis *et al.*⁴ have found some quantitative justification for this interpretation of the Au spectra by modelling the Au surface as a two-dimensional electron gas scattered by a δ -function line potential (representing a step edge). As Cu(111) supports a Shockley surface state⁹ similar to that seen on Au(111), the dI/dV spectra shown in Fig. 3 are also naturally explained by electrons tunnelling into (and out of) a surface-state band.

We now discuss why one might expect standing waves in the LDOS when surface-state electrons are scattered from step edges and point defects. We consider first the case of a single step edge. From the dI/dV spectra shown in Fig. 3 it seems that the two-dimensional surface-state electrons are strongly scattered by the step edge. We therefore use an approach similar to that of ref. 4, modelling the surface as a two-dimensional electron gas (the surface-state band) in the presence of a single hard wall barrier (the step edge) extending infinitely along the y -axis. One can write down the LDOS for such a system of independent electrons as follows:

$$\text{LDOS}(E', x) = \frac{2}{\pi^2} \sqrt{\frac{2m^*}{\hbar^2 E'}} \int_0^{k_0} dk \sqrt{\frac{1}{1 - \left(\frac{k}{k_0}\right)^2}} \sin^2(kx) \quad (1)$$

Here E' is the energy of an electron measured with respect to the surface-state band edge, x is the distance from the step edge, m^* is the effective mass of a surface-state electron, and

$$k_0 = \sqrt{\frac{2m^* E'}{\hbar^2}} \quad (2)$$

The integral in equation (1) can be solved analytically to yield

$$\text{LDOS}(E', x) = \left(1 - J_0(2k_0 x)\right) L_0 \quad (3)$$

where J_0 is the zero-order Bessel function, and $L_0 = m^*/(\pi \hbar^2)$ (L_0 is the LDOS of a two-dimensional electron gas in the absence of any scattering). This expression can be compared to dI/dV linescan data by adding a free constant offset (representing the contribution of bulk states to the tunnelling current), a free multiplicative factor (representing the effect of the STM tip's own LDOS) and a phase offset (reflecting uncertainty in the step-edge location).

Figure 2a shows fits of equation (3) to the dI/dV line scan data measured near a step edge at different bias voltages. Equation (3) provides a good fit to the decaying, oscillatory behaviour of the measured LDOS (although discrepancies do occur very close to the step edge, as one might expect for such a simple model). The values of k_0 extracted from the fit can be plotted against energy to define an experimental dispersion curve, as shown in the inset to Fig. 2a. By fitting this curve to equation (2) one can extract the surface-state effective mass ratio (m^*/m_e , where m_e is the bare electron mass) and band edge from the

LDOS spatial oscillations. This procedure yields $m^*/m_e = 0.38 \pm 0.02$ and a band edge 0.44 ± 0.01 eV below E_F , in agreement with the values obtained from photoemission studies of Cu(111)^{1,11}. The extracted value of the surface-state band edge also matches the location of the cusp in the clean terrace dI/dV spectrum of Fig. 3. The Friedel oscillation in the total surface-state density results from summing these energy-resolved WOS oscillations from the bottom of the surface-state band to the Fermi energy.

The LDOS oscillations around point defects on the Cu(111) surface can similarly be analysed in the context of a two-dimensional gas of independent electrons. If one considers a point defect as a cylindrically symmetric scattering potential, then the wavefunction of an electron in the vicinity of a point defect can be written in cylindrical coordinates as $\Psi(\rho, \phi) = R_{E'l}(\rho) e^{il\phi}$. Here $R_{E'l}(\rho)$ is the radial part of the wavefunction of an electron with energy E' and angular momentum l , and $e^{il\phi}$ is the azimuthal part of the wave function. The change in LDOS due to the presence of the point defect scatterer can then be written as

$$\Delta \text{LDOS}(E', \rho) \propto \sum_l \left(|R_{E'l}(\rho)|^2 - |J_l(k_0 \rho)|^2 \right) \quad (4)$$

Here $J_l(k_0 \rho)$ (the l th-order Bessel function) is the radial wave function of an electron in the absence of any scattering, and k_0 is defined as before¹².

For the simplest case of low-energy scattering in the asymptotic limit, the higher-order angular momentum components cancel (because of the centrifugal barrier) and the change in LDOS can be written as

$$\Delta \text{LDOS}(E', \rho) \propto \frac{1}{k_0 \rho} \left[\cos^2 \left(k_0 \rho - \frac{\pi}{4} + \partial_0 \right) - \cos^2 \left(k_0 \rho - \frac{\pi}{4} \right) \right] \quad (5)$$

Here ∂ is the phase shift of the $l=0$ partial wave. Figure 2b shows the result of fitting equation (5) to linescan data taken over a single point defect (in the same manner that equation (3) was compared to step-edge data). Equation (5) was used to fit the data only up to $3 \text{ \AA}$ from the centre of the defect. Assuming a band edge 0.44 eV below E_F , the best fit is obtained for an effective mass of $m^*/m_e = 0.38$ and an $l=0$ phase shift of $\partial_0 = -66^\circ$. Such a phase shift corresponds to a repulsive potential, in agreement with the appearance of point defects as dips on the surface.

The existence of these surface-state standing waves was, to

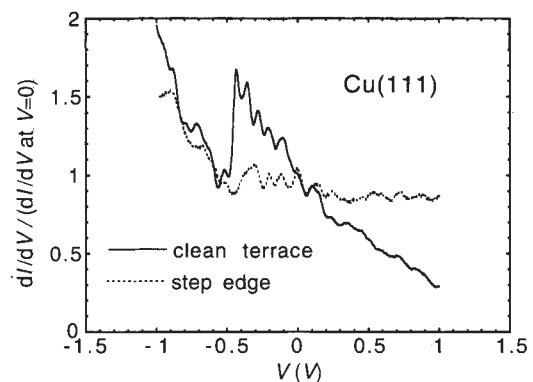


FIG. 3 dI/dV spectra (normalized at $V=0$) taken with tip over a clean terrace (solid line) and over the centre of a step edge (dashed line). The STM tip was kept stationary by opening the feedback loop during each voltage ramp (a 10-mV r.m.s., 205-Hz sine wave was used for a.c. measurement of dI/dV). The d.c. tunnel current at $+1.0$ V was 1.0 nA for both spectra. Neither raising nor lowering the d.c. tunnel current by a factor of 10 caused the spectra to change from that shown here.

some extent, anticipated in the calculations of Davis *et al.*⁴ for Au(111). Everson *et al.*⁸ did not, however, observe any manifestation of the standing waves in their STM images of the Au(111) surface at room temperature. Although we have observed that the standing waves persist at least up to 77 K, there is no indication of these standing waves: in room-temperature STM images of Cu (111) obtained elsewhere^{13,14}. The inability to image standing waves in the LDOS of Cu at room temperature may be due to the destruction of quantum phase coherence from inelastic scattering, or to instrumental limitations.

Although much can be learned from the LDOS oscillations around defects on the surface, it is also interesting to consider using known adsorbates to scatter the surface-state electrons. By using established STM adsorbate manipulation techniques⁵, it may be possible to construct nanometre-scale scattering structures to study the properties of artificially confined low-dimensional electrons.

During preparation of this manuscript we have become aware of the recent observation of similar standing waves on the Au(III) surface at room temperature (Y. Hasegawa and Ph. Avouris, unpublished results). □

Received 13 April; accepted 4 May 1993.

1. Gartland, P.O. & Slagsvold, B.J. *Phys. Rev.* **B12**, 4047–4058 (1975).
2. Heilmann, P., Neddermeyer, H. & Roloff, H.F.J. *Phys. C10*, L17–L22 (1977).
3. Zangwill, A. *Physics at Surfaces*, Ch. 4 (Cambridge Univ. Press, Cambridge, 1988).
4. Davis, L.C., Everson, M.P., Jaklevic, R.C. & Shen, W. *Phys. Rev.* **B43**, 3821–3880 (1991).
5. Eigler, D.M. & Schweizer, E.K. *Nature* **344**, 524–526 (1990).
6. Tersoff, J. & Hamann, D.R. *Phys. Rev.* **B31**, 805–813 (1985).
7. Lang, N.D. *Phys. Rev.* **B34**, 5947–5950 (1986).
8. Everson, M.P., Jaklevic, R.C. & Shen, W. *J. Vac. Sci. Technol.* **A8**, 3662–3665 (1990).
9. Shockley, W. *Phys. Rev.* **56**, 317–323 (1939).
10. Forstmann, F. *Z. Phys.* **235**, 69–74 (1970).
11. Kevan, S.D. *Phys. Rev. Lett.* **50**, 526–529 (1983).
12. Ziman, J.M. *Principles of the Theory of Solids*, Section 5.5 (Cambridge Univ. Press, Cambridge, 1972).
13. Brodde, A., Tosch, S.T. & Neddermeyer, H. *J. Microsc.* **152**, 441–448 (1988).
14. Samsavar, A. *et al.* *Phys. Rev. Lett.* **65**, 1607–1610 (1990).

ACKNOWLEDGEMENTS. We acknowledge useful discussions with V. Kalmyer and C. T. Rettner.

Molecular layering during evaporation of ultrathin liquid films

Mikel L. Forcada* & C. Mathew Mate

IBM Research Division, Almadén Research Center, 650 Harry Road, San José, California 95120-6099, USA

DYNAMICAL processes in very thin liquid films play a critical role in phenomena such as lubrication, wetting, spreading and emulsions, but these dynamics remain poorly understood at the molecular level. Heslot *et al.*^{1–5} observed molecular layering effects in a spreading liquid droplet, which they interpreted in terms of flow anomalies. Here we report that evaporation dynamics can also lead to the formation of distinct molecular layers in a thin liquid film. We use ellipsometry to follow the evolution in time of the thickness profile of a film of tetrakis(2-ethylhexoxy)silane deposited on silicon wafers. From an initially smooth gradient, we observe the formation of molecular steps 10 Å in height. The evaporation rate shows an oscillatory dependence on film thickness, leading us to conclude that layering results from faster evaporation of incomplete layers relative to complete layers.

Tetrakis(2-ethylhexoxy)silane (TEHOS) has chemical structure $[(C_4H_9)-(C_2H_5)CH-CH_2-O]_4Si$, relative molecular mass $M_R = 545$, refractive index $n = 1.439$, and density $\rho = 0.880$

$g\text{ cm}^{-3}$. The volume occupied by one molecule is $1,030\text{ Å}^3$; therefore, the thickness of a molecular layer should be $\sim 10\text{ Å}$. The smooth Si(100) silicon wafers used were 25.4 mm in diameter, polished on one side, and had a native silicon oxide layer 12–16 Å thick. After cleaning the wafers with 2-propanol and 1,1,2-trichlorotrifluoroethane, we deposited films by withdrawing the wafer from a solution of TEHOS in 1,1,2-trichlorotrifluoroethane solvent. This volatile solvent evaporates completely within a few seconds after withdrawal.

Immediately after deposition, the initial thickness profile of each film was measured by ellipsometry along the withdrawal direction. The ellipsometer spot size was 1 mm along the direction of the profile and 2.5 mm perpendicular to the profile. Each wafer was then stored horizontally in a covered glass Petri dish and placed in a room-temperature ($24 \pm 2^\circ\text{C}$) enclosure with a continuous flow of dry, filtered nitrogen gas. The wafers were exposed to laboratory air only during subsequent ellipsometric measurements. Adsorption of contaminants on clean control wafers stored alone in the Petri dishes was insignificant (less than 1 Å in 20 days).

Figure 1 shows two examples of the time evolution of TEHOS films with a thickness gradient created by varying the withdrawal speed during coating. In Fig. 1a, the initially smooth gradient across the wafer evolves over the first 187 h to a flat terrace, 38 Å or four molecular layers thick, with a 10 Å step down to a short terrace, 28 Å or three molecular layers thick. The area underneath the thickness profile decreases with time, suggesting that the film is slowly evaporating. If flow were the dominant process, the profile would shift to the right because of the gradient in the disjoining pressure¹, while becoming more level

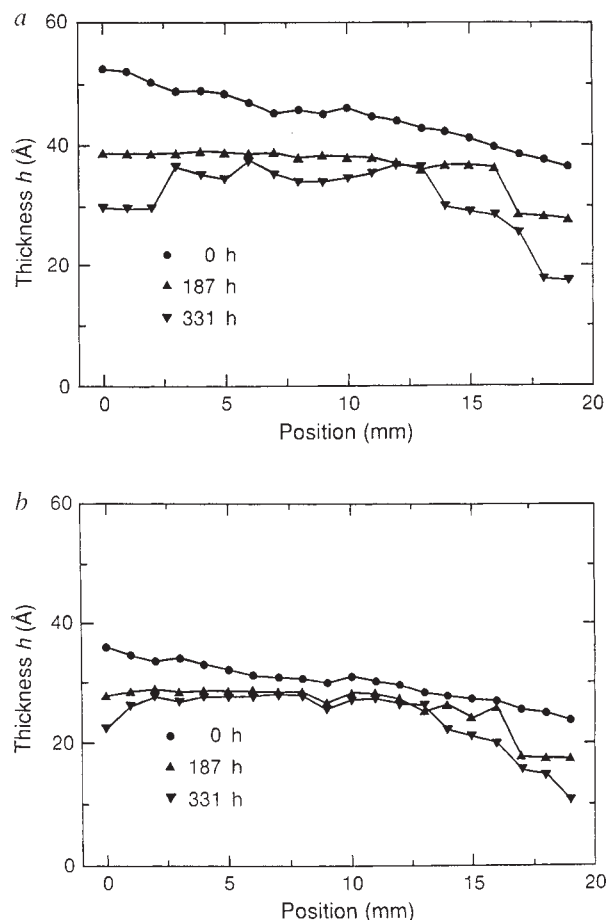


FIG. 1 Ellipsometric thickness profiles, measured over the 20 mm centre portion of the wafer, of two non-uniform TEHOS films deposited on silicon wafers. The profiles change with time in both cases.

* Permanent address: Departament de Física Aplicada, Universitat d'Alacant, E-03071 Alacant, Spain.

so as to conserve area underneath the profile. Flow to the underside of the wafer is unlikely as our film preparation method deposits a similar profile of TEHOS on that side. The influence of gravity on flow is estimated to be at least 10^5 times less important than that of disjoining pressure. Figure 1b shows a rather thinner initial smooth gradient profile, which again evolves over the first 187 h to form two flat terraces, two and three molecular layers thick. From both Fig. 1a and b, it is apparent that once a terrace of a complete multilayer forms the thinning rate slows down considerably.

To investigate further the importance of flow, we prepared several wafers with the TEHOS film removed from a portion of the surface, by partially dipping a uniformly coated wafer in pure 1,1,3-trichlorotrifluoroethane solvent. These wafers were stored supported above the surface of the Petri dish to minimize the possibility of TEHOS flow from the wafer to the glass surface. Figure 2 shows a typical example of the time evolution of a TEHOS film with the right-hand portion removed. The film slowly thins over time, with distinct 10-Å-high steps becoming apparent after 288 h. If flow were the dominant thinning process, TEHOS would flow rapidly from left to right because of the steep thickness gradient and the correspondingly large gradient in disjoining pressure¹. But the area gained on the right-hand side of the profiles is only 10% of the area lost underneath the profiles on the left-hand side, indicating that most of the thinning of the film is due to mass loss rather than flow. Again, flow onto the underside of the wafer is unlikely as the film profile there is similar to the top surface. In Fig. 2, the lack of significant flow left to right, together with the improbability of flow off the left edge of the wafer, indicates that evaporation, not flow, is the dominant mechanism for the observed thinning.

The relationship between evaporation and the formation of distinct molecular layers can be seen more clearly by measuring the thinning rate as a function of film thickness. Average thinning rates are determined as follows. Let $h(t)$ be the thickness at a given position and time t , and let t_1 and t_2 be the times of two consecutive profiles. An approximate thinning rate $-h(t_2) - h(t_1)/(t_2 - t_1)$ is then assigned to the thickness interval $[h(t_2), h(t_1)]$, typically much narrower than a monolayer. This calculation is repeated for all positions and times, and the approximate thinning rates are averaged where thickness intervals overlap. Figure 3 shows the results obtained for 1,060 positions and times on five wafers with initially smooth gradient profiles. The thinning rate $-dh/dt$ oscillates, with minima roughly every 10 Å, the expected molecular layer thickness, and

shows that complete molecular layers evaporate more slowly than incomplete multilayers. The method of averaging over intervals, which also averages out uncertainties from the thickness measurement process, tends to smooth the final result, overestimating the evaporation rate of complete multilayers and underestimating the rate for incomplete multilayers. Because of the averaging, the faster thinning at the wafer edges visible in Figs 1 and 2 does not appear in the computed result for $-dh/dt$. The locations of the first two minima, 6 and 16 Å, are substantially lower than the 10 and 20 Å expected for one and two molecular layers, suggesting that the molecules in these first two layers may be flattened compared with those in thicker films, although we cannot rule out a systematic error in the measurement of the thickness of these thinner films.

The oscillatory behaviour of the thinning or evaporation rate can be understood from the difference in the heat of desorption for molecules in complete layers compared with those in incomplete layers. Molecules in incomplete layers have on average fewer neighbouring molecules, leading to a smaller value of the heat of desorption. Indeed, for the adsorption of gases on surfaces where a step-like adsorption isotherm is observed, the heat of desorption is typically found to have an oscillatory dependence on coverage with a maximum occurring at the completion of each layer⁶. Molecules within an incomplete layer would find it energetically favourable to coalesce to form islands, with evaporation occurring preferentially at the edges. Indeed, in Figs 1 and 2, thinning seems to occur preferentially at step edges or at the wafer edges. The stepwise condensation and redissolution of crystal surfaces, described in, for example, ref. 7, may be a closely related process.

Heslot and colleagues¹⁻⁵ have reported the formation of molecular layers by spreading droplets of TEHOS and polydimethylsiloxane, and for TEHOS in a capillary-rise geometry. Several authors using analytical^{5,8,9} and molecular simulation¹⁰⁻¹² methods have explored possible dynamical flow mechanisms for the formation of these molecular layers. However, the oscillatory nature of the evaporation rate shown in our experiments leads us to reinterpret the layering observed for spreading droplets of this liquid. During the early stages of spreading, the droplet slowly evaporates. Once the thin film extending from the spreading droplet evaporates down to complete molecular layers, evaporation slows considerably, stabilizing the formation of the molecular layers. The slower evaporation rate of complete molecular layers within the droplet would make it appear that the volume of the spreading droplet was conserved,

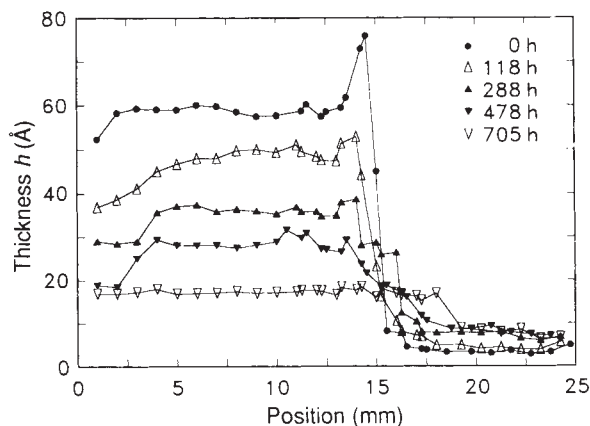


FIG. 2 Time evolution of the ellipsometric thickness profiles of a TEHOS film deposited on a silicon wafer, after the right-hand portion of the film has been removed.

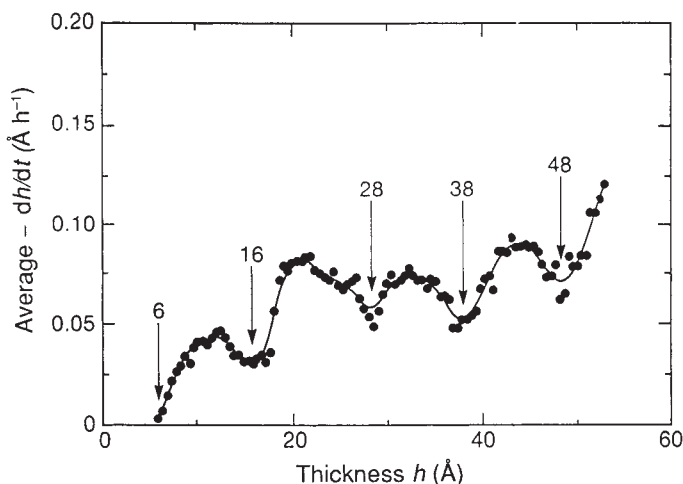


FIG. 3 Average thinning rate $-dh/dt$ of a set of five TEHOS films as a function of film thickness h . The solid line is a smoothing-spline curve with a r.m.s. deviation of 0.004 Å h^{-1} .

resulting in the erroneous conclusion that the liquid was nonvolatile.

Our results also bring into question whether the spreading is mainly from flow or diffusive processes as previously assumed^{15,18-10}. We have found that if an initially clean wafer is placed several millimetres away from a TEHOS-coated wafer in the same Petri dish, the clean wafer will develop a 5-Å-thick film over a few weeks, by readsorbing TEHOS that has evaporated from the coated wafer. This result indicates the importance of evaporation and readsorption to the spreading of TEHOS, recently also shown to be the dominant spreading mechanism for droplets of moderately volatile liquids¹³. An unresolved question is whether molecular layering can occur in other types of liquid films solely by flow, as originally suggested⁴. By experimenting with liquids whose volatility relative to their ability to flow is much lower than that of TEHOS, it should be possible to distinguish between evaporation and flow mechanisms in the molecular layering. □

Received 13 October 1992; accepted 20 April 1993.

1. Cazabat, A. M., Fraysse, N., Heslot, F. & Carles, P. *J. phys. Chem.* **94**, 7581-7585 (1990).
2. Heslot, F., Fraysse, N. & Cazabat, A. M. *Nature* **338**, 640-642 (1989).
3. Heslot, F., Cazabat, A. M. & Levinson, P. *Phys. Rev. Lett.* **62**, 1286-1289 (1989).
4. Heslot, F., Cazabat, A. M., Levinson, P. & Fraysse, N. *Phys. Rev. Lett.* **65**, 599-602 (1990).
5. Cazabat, A. M., Heslot, F. & Fraysse, N. *Prog. Colloid Polymer Sci.* **83**, 52-55 (1990).
6. Prenzl, C. F. & Halsey, G. D. *J. phys. Chem.* **61**, 1158-1165 (1957).
7. Hillner, P. E., Manne, S., Gratz, A. J. & Hansma, P. K. *Ultramicroscopy* **42-44**, 1387-1393 (1992).
8. De Gennes, P. G. & Cazabat, A. M. *C.R. Acad. Sci. Paris, Ser. II* **310**, 1601-1606 (1990).
9. Fraysse, N., Cazabat, M. & Cazabat, A. M. *C.R. Acad. Sci. Paris, Ser. II* **314**, 1025-1028 (1992).
10. Nieminen, J. A., Abraham, D. B., Karttunen, M. & Kaski, M. *Phys. Rev. Lett.* **69**, 124-127 (1992).
11. Yang, J.-x., Koplik, J. & Banavar, J. R. *Phys. Rev. Lett.* **67**, 3539-3542 (1991).
12. Yang, J.-x., Koplik, J. & Banavar, J. R. *Phys. Rev.* **A46**, 7738-7749 (1992).
13. Novotny, V. J. & Marmur, A. *J. Colloid Interf. Sci.* **145**, 355-361 (1991).

ACKNOWLEDGEMENTS. We thank G. Vurens, A. Homola, V. Novotny and C. Gudeman for useful discussions, and the Ministry of Education and Science of Spain and the Fulbright Commission in the U.S. for post-doctoral support for M.L.F.

Understanding the acid behaviour of zeolites from theory and experiment

G. J. Kramer, R. A. van Santen, C. A. Emeis & A. K. Nowak

Koninklijke/Shell-Laboratorium, Amsterdam, (Shell Research B. V.), PO Box 3003, 1003 AA Amsterdam, The Netherlands

ZEOLITES are microporous aluminosilicates which, in their protonated form, act as solid catalysts¹, and are widely used in the oil and petrochemical industries for processes such as cracking, isomerization and alkylation of hydrocarbons². The proposed mechanisms³⁻⁵ of these processes mostly involve proton transfer and formation of carbenium or carbonium ions as reactive intermediates, but the detailed function of the zeolite and in particular the relation between acidity and catalytic activity is not well understood. Here we report experimental and theoretical studies of deuterium-hydrogen exchange between deuterated methane and protonated zeolites — a prototypical heterogeneous catalytic reaction between a hydrocarbon and an acid zeolite. We monitored this slow exchange reaction in two different zeolites using infrared spectroscopy, and used *ab initio* quantum chemistry calculations to determine both the reaction mechanism and the acidity-activity relationship. Combining our

theoretical results with recent estimates⁸⁻¹¹ of the acidity differences within zeolites enables us to reproduce the experimentally observed reaction rates and thus to obtain a detailed microscopic picture of this heterogeneous catalytic process.

The isotope exchange reaction was monitored by measuring the evolution of the infrared absorption spectrum of a zeolite disk in a closed container in contact with an excess of deuteromethane. Two zeolite samples were used, one with the faujasite (FAU) structure (Si:Al ratio = 10), the other having the ZSM-5 (or MFI) structure (Si:Al ratio = 13). Over time, we observed the disappearance of the absorption bands corresponding to the OH stretch vibration of acid Al-O(H)-Si and terminal SiOH (silanol) groups, with the simultaneous appearance of the corresponding OD and SiOD bands. A control experiment with the pure-silica form of MFI, which contains no acid OH groups, showed that silanol groups are not active in the exchange process. The simultaneous replacement of silanol and acid hydrogen by deuterium in the acid zeolites should be attributed to the intra-zeolite exchange of hydrogen, possibly assisted by trace amounts of water, which is rapid compared with the D/H exchange between the zeolite and deuteromethane^{6,7}. The rate of conversion of acid OH groups was measured at temperatures between 620 and 750 K for a methane gas pressure of 13 torr. The lower temperature limit was dictated by the low reaction rate and the upper limit by the onset of methane pyrolysis. The experimental exchange rates (Fig. 1) show a clear difference in activity between FAU and MFI.

We will now turn to first-principles calculation of the results. We apply quantum chemistry to a molecular system in which the zeolite is represented by a molecular cluster. The exchange process is influenced in real zeolites by the variation of proton affinity between crystallographically different oxygen sites. For FAU and MFI, these differences have been quantified⁸⁻¹¹. We will show that proton affinity differences alter that rate of exchange enough to account for the observed activity difference between zeolites FAU and MFI.

The molecular system used in the *ab initio* quantum chemical study consisted of methane and a tri-tetrahedral SiH₃-OH-Al(OH)₂-O-SiH₃ cluster representing the acid zeolite. The peripheral bonds of the cluster, which are in reality connected to the zeolite framework, were saturated with hydrogen atoms. Because of the flexibility of the zeolite lattice, such a cluster is a good model system for the extended

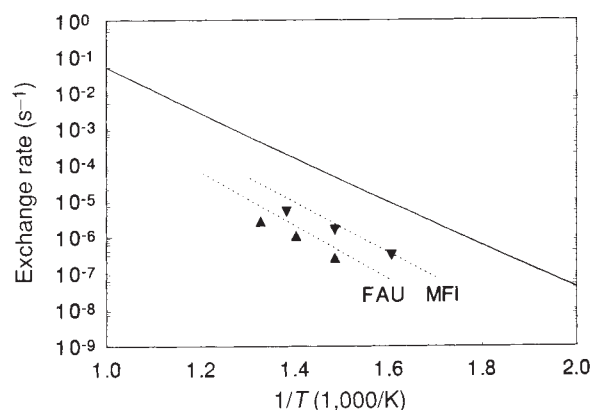


FIG. 1 D/H exchange rate of deuterated methane with the H-forms of zeolites FAU and MFI. Experimental values are given by the full symbols (FAU ▲, MFI ▼). The full line is the prediction from transition-state theory for the exchange reaction with a zeolite cluster, which has an effective reaction barrier of 122 kJ mol⁻¹. Proton affinity differences reduce the exchange rate in real zeolites. The theoretical predictions for zeolites FAU and MFI are indicated by dashed lines.

zeolite^{12,13}. Both the adsorption minimum of methane and the transition state for D/H exchange were found using standard conjugate-gradient geometry-optimization methods¹⁴. The energy difference between the adsorption minimum and transition state for self-consistent-field (SCF) calculations¹⁵ is 180 kJ mol⁻¹ when using a 3-21G basis and 230 kJ mol⁻¹ with a larger 6-31G** basis. Additional single-double configuration interaction calculations including Pople size-consistency corrections¹⁵ allowed for the estimation of electron correlation effects which are neglected at the SCF level. Electron correlation reduces the barrier by 40 to 60 kJ mol⁻¹, leading to an estimate of 150 ± 20 kJ mol⁻¹ for the barrier. Figure 2 shows the calculated transition state and the reaction coordinate.

The reaction coordinate represents the transfer of the proton of the zeolite to the methane molecule and the symmetrical return of one of the deuterium atoms to the zeolite. Thus, the exchange process uses two of the oxygen atoms of the zeolite: one as an acid, the other as a base. In the transition state the exchanging atoms are halfway between the oxygen and carbon atoms, at distances of 1.201 and 1.391 Å, respectively. These short distances indicate that, in the transition state, the hydrocarbon fragment is covalently bonded to the zeolite lattice.

Once the barrier height and the results of the vibrational analysis of initial state and transition state are known, the exchange rate can be theoretically predicted using Eyring's transition state theory¹⁶. This approach has been successfully applied for simple gas-phase reactions^{17,18}. As measurements were made under constant pressure, the reference value for the (free) energy is that of the gas phase. Hence the effective, experimental barrier, on which the exchange rate depends exponentially, is equal to the barrier as calculated from the cluster calculations, corrected by the adsorption energy of methane in the zeolite at zero temperature, which is ~ -28 kJ mol⁻¹ for both FAU and MFI¹⁹⁻²¹. The pre-exponential factor depends on the entropy change between the initial state and the transition state, determined by the respective vibrational frequencies. Quantum corrections²² turned out to be of minor importance. The resulting prediction is shown in Fig. 1, represented by a solid line.

So far we have treated the exchange process as if the zeolite were a molecular catalyst. In doing so, we have neglected the variation of acidity within the zeolite lattice. Recent theoretical studies^{8,11} using classical force-field methods, have estimated these differences in zeolites FAU and MFI to be as large as 60 kJ

mol⁻¹. The exchange rate obviously changes substantially if proton affinity differences affect the exchange barrier.

We established the relation between acidity and catalytic activity for the D/H exchange reaction through a series of quantum-chemical cluster calculations in which the peripheral SiH bonds of the cluster were fixed at varying distances. As a result, the strength of the SiO bond, and thereby the strength of the acid OH bond, could be strengthened or weakened at will. On changing the SiH distances of one of the peripheral SiH₃ groups between 1.3 and 1.7 Å, the proton affinity of the associated oxygen atom varied by ~ 60 kJ mol⁻¹, thereby mimicking the acidity variation in real zeolites.

Two series of calculations were done. First, the SiH distances in the two SiH₃ groups were varied jointly. Thus, the proton binding strength (acidity) of both oxygen atoms involved in the exchange process remained equal. The barrier height was unaffected, which is explained by the covalent character of the transition state in which the OH bond strength of the initial state is shared between two partial OH bonds (Fig. 2). Second, we simulated the effect of acidity differences by lowering the proton affinity of the proton-accepting oxygen site only. In this case, again because of the covalent bond-sharing character of the transition state, the barrier was increased, proportional to the difference in proton affinity between donor and acceptor site (with constant of proportionality 0.7). In summary, the covalent character of the exchange process enforces a non classical relationship between acidity and activity, where activity is determined by acidity differences between the proton donating and accepting oxygen sites.

As the reaction rate is the product of occupation by protons of a given oxygen site and the probability of exchange from that site, it is always adversely affected by proton affinity differences. Therefore the exchange rate found for the molecular catalyst, where proton affinities are equal, is predicted to be the upper limit for the exchange rate in any aluminosilicate material.

Having established the relation between proton affinity differences and the rate of D/H exchange, we may use the proton affinity data for FAU and MFI from ref. 11 to calculate the reduction of the exchange rate in these zeolites. For both zeolites all T-sites which are the sites that can be occupied by aluminium, were considered. For each of these there are six possible exchange paths between any of the four oxygen atoms surrounding the central aluminium. Some paths are sterically inaccessible: in zeolite FAU, for instance, there is a unique crystallographic T-site²³ surrounded by four difference oxygen

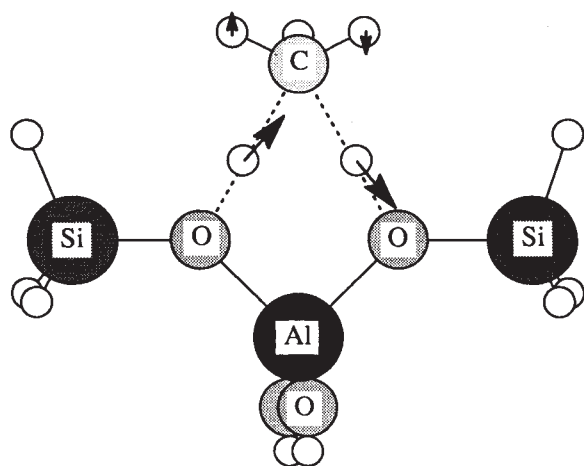


FIG. 2 Calculated transition-state geometry for the isotope exchange process between methane and an acid zeolite cluster. All small circles represent hydrogen or deuterium atoms. The arrows indicate the main components of the displacement vectors along the reaction coordinate.

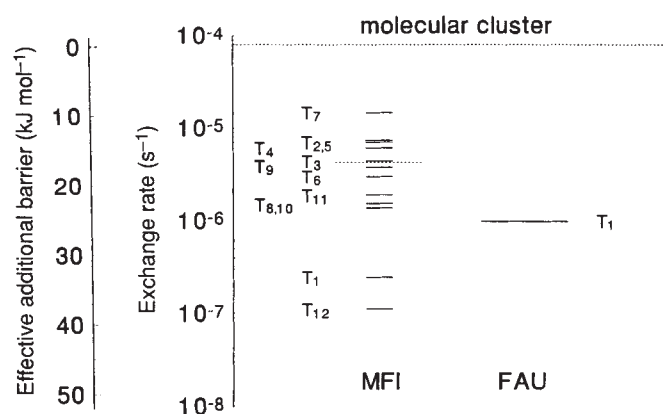


FIG. 3 Site-dependent exchange rates in zeolites FAU and MFI at 700 K and 13 torr methane pressure. These zeolites have a lower exchange rate than the molecular catalyst, because a higher effective barrier is induced by proton affinity differences between the oxygen sites. The dashed line indicates the average exchange rate for MFI. The numbering of T-sites in zeolite MFI is taken from ref. 23; FAU has one unique T-site.

sites, one of which does not take part in the exchange process because a proton on this site protrudes into the sodalite cage, which is inaccessible to methane. For the remaining three accessible pathways, the proton affinity differences are 35 kJ mol⁻¹ or higher. In combination with the appropriate Boltzmann factors for the proton occupancy of oxygen sites, this leads to a reduction of the exchange rate by a factor of 80 at 700 K compared with the molecular catalyst.

MFI has a more complex crystal structure with 12 different T-sites²³. For each of these, we repeated the calculation outlined above. Between different T-sites, the exchange rate varied by two orders of magnitude (Fig. 3). The large activity differences corroborate empirical evidence that the acid sites are heterogeneous²⁴. Assuming a random distribution of aluminium over all T-sites, the overall exchange rate in MFI is the average of these exchange rates, which at 700 K is a factor of 20 lower than that of the molecular catalyst and thus a factor 4 larger than in FAU. These reduction factors vary with temperature; the predicted reaction-rate dependencies are shown by dashed lines in Fig. 1.

The theoretical predictions of the relative rates for the two zeolites and their temperature dependence agree well with experiment. It should be mentioned that there is a 20 kJ mol⁻¹ uncertainty in the absolute value of the barrier as predicted by quantum chemistry. But this uncertainty affects the predictions for all zeolites to the same extent, so the prediction of a factor of four difference in reactivity between zeolites FAU and MFI is not affected.

Our results show that heterogeneous catalysts can be understood at the atomic level, and that quantitative agreement between theory and experiment can be achieved. For the exchange reaction studied here, we have predicted the reaction path and established the relationship between acidity and catalytic activity. When combined with theoretical estimates for the proton affinity of the various oxygen sites within the zeolite, this gives an *ab initio* prediction of heterogeneous catalytic activity. Applying the argument in reverse, it gives a first, albeit indirect, experimental measure of the acidity differences in zeolites, which must be ~ 60 kJ mol⁻¹ to explain satisfactorily the reactivity differences between FAU and MFI. □

Received 29 January; accepted 7 May 1993.

1. Thomas, J. M. *Scient. Am.*, **266** (IV), 82–88 (1992).
2. Maxwell, I. E. & Stork, W. H. J. in *Introduction to Zeolite Science and Practice*, (eds van Bekkum, H. et al.) 571–630 (Elsevier, Amsterdam, 1991).
3. Jacobs, P. A. & Martens, J. A. in *Introduction to Zeolite Science and Practice*, (eds van Bekkum, H. et al.) 445–496 (Elsevier, Amsterdam, 1991).
4. Abbot, J. & Wojciechowski, B. W. *J. Catal.* **115**, 1–15 (1989).
5. Stefanidis, C., Gates, B. C. & Haag, W. O. *J. Molec. Catal.* **67**, 363–367 (1991).
6. Sauer, J., Kölmel, C. M., Hill, J. R. & Ahlrichs, R. *Chem. Phys. Lett.* **164**, 193–198 (1989).
7. Sauer, J., Horn, H., Häser, M. & Ahlrichs, R. *Chem. Phys. Lett.* **173**, 26–32 (1990).
8. Dubsky, J., Beran, S. & Bosáček, V. *J. Molec. Catal.* **6**, 321–326 (1979).
9. Schröder, K.-P., Sauer, J., Leslie, M., Catlow, R. C. A. & Thomas, J. M. *Chem. Phys. Lett.* **188**, 320–325 (1992).
10. Schröder, K.-P., Sauer, J., Leslie, M. & Catlow, R. C. A. *Zeolites* **12**, 20–23 (1992).
11. Kramer, G. J. & van Santen, R. A. *J. Am. chem. Soc.* **115**, 2887–2897 (1993).
12. Kramer, G. J., de Man, A. J. M. & van Santen, R. A. *J. Am. chem. Soc.* **113**, 6435–6441 (1991).
13. Sauer, J. *chem. Rev.* **89**, 199–255 (1989).
14. Dupuis, M., Sprangler, D. & Wendolowski, D. *NRCC Software Catalog 1*, Program No. QG01, GAMESS (1980).
15. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Macmillan, New York, 1982).
16. Rice, O. K. *Statistical Mechanics, Thermodynamics and Kinetics* (Freeman, San Francisco, 1966).
17. Thruhar, D. G. & Gordon, M. S. *Science* **249**, 491–497 (1990).
18. Gilbert, R. G. & Smith, S. C. *Theory of Unimolecular and Recombination Reactions* (Blackwell, Oxford, 1990).
19. Barrer, R. M. & Sutherland, J. W. *Proc. R. Soc. A* **237**, 439–450 (1956).
20. Papp, H., Hinsen, W., Do, N. T. & Baerns, M., *Thermochim. Acta* **82**, 137–148 (1984).
21. Yashonath, S., Thomas, J. M., Nowak, A. K. & Cheetham, A. K. *Nature* **331**, 601–604 (1988).
22. Johnston, H. S. *Gas Phase Reaction Rate Theory* (Ronald, New York, 1966).
23. Meier, W. M. & Olson, D. H. *Atlas of Zeolite Structure Types*, 2nd edn (Butterworth, Cambridge, 1987).
24. Lombardo, E. A., Sill, G. A. & Hall, W. K. *J. Catal.* **119**, 426–440 (1989).

ACKNOWLEDGEMENTS: We thank B. J. Milan for help in the experimental work.

Milankovitch theory viewed from Devils Hole

J. Imbrie*, A. C. Mix† & D. G. Martinson‡

*Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

†College of Oceanography, Oregon State University, Corvallis, Oregon, 97331, USA

‡Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

VARIATIONS in the oxygen isotope content ($\delta^{18}\text{O}$) of late Quaternary deep-sea sediments mainly reflect changes in continental ice mass¹, and hence provide important information about the timing of past ice ages. Because these sediments cannot yet be dated directly beyond the range of radiocarbon dating (40–50 kyr), ages for the $\delta^{18}\text{O}$ record have been generated^{2,3} by matching the phase of the changes in $\delta^{18}\text{O}$ to that of variations in the Earth's precession and obliquity. Adopting this timescale yields a close correspondence between the time-varying amplitudes of these orbital variations and those of a wide range of climate proxies⁴, lending support to the Milankovitch theory that the Earth's glacial-interglacial cycles are driven by orbital variations. Recently Winograd *et al.*⁵ reported a record of $\delta^{18}\text{O}$ variations in a fresh-water carbonate sequence from Devils Hole, Nevada, dated by U–Th disequilibrium⁶. They concluded that the timing of several of the features in the record, which reflects changes in the temperature of precipitation over Nevada as well as changes in the isotopic composition of the moisture source^{5,7}, showed significant deviations from that predicted by Milankovitch theory. Here we demonstrate that applying the Devils Hole chronology to ocean cores requires physically implausible changes in sedimentation rate. Moreover, spectral analysis of the Devils Hole record shows clear evidence of orbital influence. We therefore conclude that transfer of the Devils Hole chronology to the marine record is inappropriate, and that the evidence in favour of Milankovitch theory remains strong.

The Devils Hole $\delta^{18}\text{O}$ record was obtained from a 36-cm core (DH-11) drilled in a layer of calcite lining a water-filled cavern in southern Nevada. Ground water here originates as snow and rain in the surrounding mountains⁵. The time resolution of the $\delta^{18}\text{O}$ record is comparable to that in deep-sea records and contains 21 precise radiometric dates⁶ from the ²³⁰Th method⁸. The general resemblance of the Devils Hole $\delta^{18}\text{O}$ record to that of the ocean is striking (Fig. 1). Indeed, the coherency is so high that it is tempting to overlook how different the physics underlying each record must be and jump to the conclusion that they are in fact synchronous. The SPECMAP marine $\delta^{18}\text{O}$ stack is an average of many open-ocean records in which values become heavier during glacial intervals, mainly because of storage of fresh water as ice. In contrast, the $\delta^{18}\text{O}$ data in DH-11 reflect the isotopic distillation of atmospheric moisture, a process linked to local precipitation temperature⁵. Values in this record therefore become lighter during glacial intervals. But to interpret these data properly as an air temperature signal, one must make assumptions not only about the isotopic composition of the oceanic moisture source (which changes over a glacial cycle), but also about air-parcel trajectories at the relevant seasons⁷. Winograd *et al.*⁵ address some, but not all, of these complex issues.

A convenient way of comparing the DH-11 and SPECMAP chronologies is to examine a mapping function⁹ that adjusts the ages in one curve to give the best match with $\delta^{18}\text{O}$ variations in the other (Fig. 2). Overall, the similarity between these physically very different records is remarkable ($r=0.85$). But we focus here on two glacial intervals, stages 6 and 10, in which the DH-11 ages are significantly older.

We consider three possible explanations for these differences: (1) that the $\delta^{18}\text{O}$ events recorded in these groundwater and oceanic records are synchronous, and the DH-11 chronology is wrong; (2) that the events are synchronous, and the SPECMAP chronology is wrong; or (3) that the events are not synchronous, and both chronologies are right. The first possibility has been examined in two recent papers that ask if Devils Hole ages could be biased by the incorporation of ^{230}Th during calcite growth^{10,11}. Although some avenues of research remain to be explored, in the absence of compelling evidence for such a bias we will assume that there is nothing wrong with the chronology of DH-11.

To test the second possibility, we examine the geological consequences of applying the DH-11 chronology to the oceanic $\delta^{18}\text{O}$ record (Fig. 1b). We do this by transferring the Devils Hole chronology to a set of four deep-sea cores (Fig. 3). The records chosen come from widely separated locations away from continental margins, at sites (in the eastern¹² and western¹³ tropical Pacific, equatorial Atlantic¹³, and Subantarctic¹⁴ oceans) where sediments accumulate at different average rates, in very different oceanographic settings and with different histories of surface and deep-water masses. Radiocarbon studies constrain models of sedimentation across the stage 2–1 transition at these sites. For V28-238 and V30-40, published dates are available^{15,16}. For the other two cores there are nearby records with accelerator mass spectrometry ^{14}C dating^{17,18}. During late glacial time, sedimentation rates (converted to calendar years) were 30% lower than post-glacial rates in V28-238; 40% higher near RC13-110; a factor of 5 lower near RC11-120; and essentially unchanged in V30-40.

The one thing we would not expect to find during glacial maxima is a synchronous increase in sedimentation rates at all

four sites. Yet this is what application of the DH-11 chronology implies for stage 6. Large anomalies in sedimentation rates in stage 6 are produced in all four cores, with anomalously high rates followed immediately by anomalously low rates (varying by a factor of 5 or 6). No similar excursions measured at this resolution have been detected in these cores over the past 300 kyr on either chronology, and these cores lack lithological changes that would be expected to accompany such excursions. In contrast, sedimentation-rate variations implied by the SPECMAP age model are small, spread rather evenly through the depth of each core, and are not similar at all sites. Similar arguments apply to the stage 10 anomaly.

We therefore reject the second hypothesis (that $\delta^{18}\text{O}$ events in DH-11 and the ocean are synchronous) and reaffirm our earlier conclusion, based on detailed investigations of relative accumulation rates^{19,20}, that the SPECMAP chronology (ref. 2, with modification in the older section in refs 21–23, and higher resolution provided in ref. 3) is a suitable age framework for Quaternary deep-sea sediments. Certainly the absolute chronology of this common temporal framework can be improved²⁴. For the stage 6–5 transition, we expect that improvements will come from dates on unaltered corals tied directly to early portions of the sea-level rise in stage 6 (ref. 25) rather than from dates, however accurate they may be, that fix the timing of events on land. As yet, such dates are not available. Meanwhile, we note that the most reliable coral dates (those with modern values of initial $^{234}\text{U}/^{238}\text{U}$ on essentially pure aragonite samples of the reef-crest coral *Acropora palmata*) for the highest stand of sea level in Barbados, Haiti and the Bahamas (122 to 126 kyr before present) seem to support this part of the SPECMAP chronology^{25,26}.

For the present, therefore, we will assume that both the SPECMAP and the Devils Hole chronologies are right when applied to their respective records, and we ask whether the Devils Hole record, displayed on its own timescale, is “inconsistent with the Milankovitch hypothesis that orbitally controlled variation in solar insolation plays a direct role in Pleistocene climate change”²⁵. To address this problem convincingly in the time domain, we require two models that are not yet available: (1) a time-dependent model of processes governing the response of the climate system to Milankovitch forcing, including those in

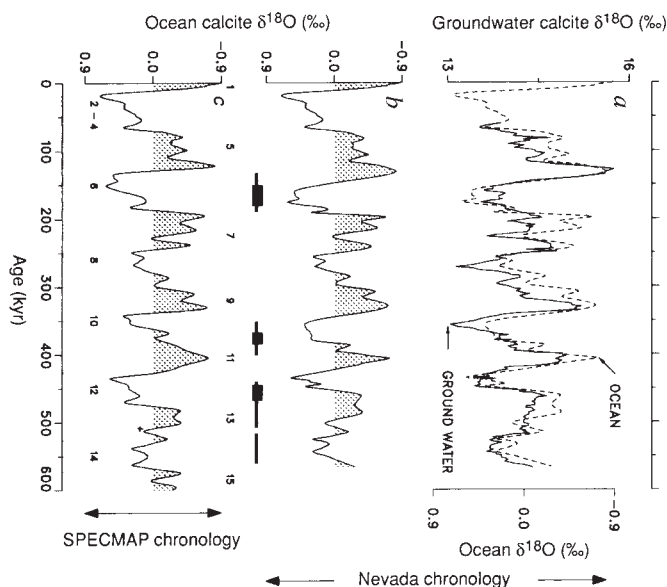


FIG. 1 A stacked $\delta^{18}\text{O}$ record² from ocean calcite (a, b, c) compared with the $\delta^{18}\text{O}$ record from groundwater calcite at Devils Hole, Nevada⁵ (a) on two different chronologies. No effort is made to eliminate the temperature effect in the oceanic ($\delta^{18}\text{O}$) record of sea level, or the $\delta^{18}\text{O}$ moisture-source effect in the groundwater record of atmospheric temperature⁷. In a and b, the SPECMAP oceanic chronology (c) has been adjusted⁹ to maximize the correlation between the two records. Shaded portions designate interglacial stages in the standard sequence of numbered marine $\delta^{18}\text{O}$ stages²⁰. The stage 6–5 transition in the DH-11 chronology (b) occurs much more slowly, relative to the 2–1 transition, than in the SPECMAP chronology (a pattern that conflicts with inferences drawn from multiple sediment records normalized in the depth domain to obtain a relative chronology^{19,20}). Black bars show where the two chronologies differ by more than 10 kyr (narrow bars) and 20 kyr (wide bars).

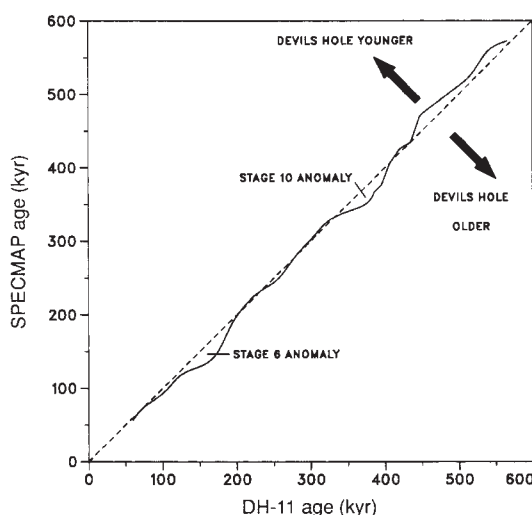


FIG. 2 Mapping function⁹ used to optimize the correlation between the oceanic and groundwater $\delta^{18}\text{O}$ curves in Fig. 1a. Mapping functions and the oceanic data used in this paper are available in digital form as SPECMAP Archives 2 and 3 at the National Center for Atmospheric Research and at the National Geophysical Data Center, both in Boulder, Colorado.

the atmosphere over Nevada and in ground waters of the Great Basin; and (2) a model that will isolate the local temperature component of the DH-11 record. But knowing that large ice sheets have time constants of the order of 15 kyr (refs 27, 28), we can still make one confident prediction: during the transition to an interglacial state, some part of the atmosphere must warm substantially before ice sheets melt and lower the fraction of ^{18}O in the ocean.

Lacking the desired models, we can make some progress by expressing the data as a spectrum (Fig. 4), thus breaking a difficult problem into simpler parts. In particular, we can look for evidence of linear responses to orbital forcing near 23 and 41

kyr (ref. 4), where virtually all the insolation change occurs, and consider these mechanisms separately from those which drive the broader-band, 100-kyr cycle (in which the response to any forcing, whether internal or external, must involve strong nonlinearities). As in the marine data, the spectrum of the Devils Hole data reveals three main concentrations of variance (Fig. 4a). Could these reflect orbital influence? Peaks in the DH-11 spectrum occur at 91 kyr and 42 kyr, with two adjacent peaks at 27 kyr and 23 kyr, rather than in bands centred at the periods of the orbital forcing (100 kyr, 40 kyr and 23 kyr) during this interval as in the marine spectrum. Still, considering the bandwidth of our analysis, the existence of discrete, coherent concentrations of variance at or near orbital periods in both records suggests orbital influence. This suggestion seems to be confirmed by the observation that the DH-11 record is strongly coherent ($r > 0.94$) with the two dominant cycles of insolation at 23 kyr and 40 kyr (Fig. 4b). The peak near 27 kyr could reflect the same nonlinear interaction tone between orbital frequencies (~ 29 kyr) that is found in records of the trade winds^{29,30}.

Thus even when we assume that the Devils Hole record is a well-dated and simple measure of climate, we find evidence of orbital influence. We therefore reject the notion that this record as a whole is inconsistent with modern versions of the Milankovitch theory (see, for example refs 31–33). Looking beyond Devils Hole to other parts of the late Pleistocene world⁴, and much farther back in time^{21, 23, 34–37}, we find compelling evidence in favour of the Milankovitch theory. □

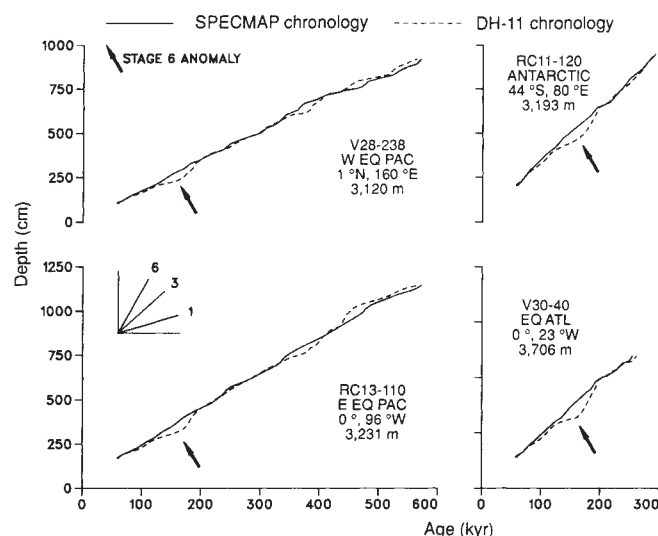


FIG. 3 Depth against age in four deep-sea cores^{1,12–14} according to the oceanic (SPECMAP) and Nevada (DH-11) chronologies. Slopes in the insert (lower left) show accumulation rates in cm kyr^{-1} . Large and physically implausible changes in sedimentation rate in stages 6 and 10 are implied by the application of the DH chronology to these cores.

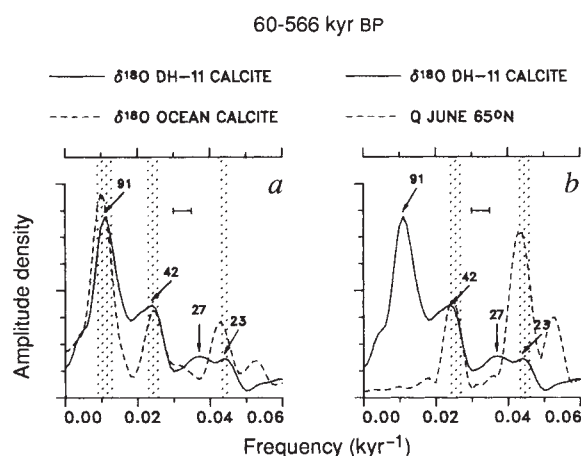


FIG. 4 Spectra of $\delta^{18}\text{O}$ records in calcite from DH-11 (Nevada) and the ocean calculated^{4,38} at a bandwidth of 0.005 kyr^{-1} on the assumption that the chronology developed for each record is correct. a, Comparison of $\delta^{18}\text{O}$ spectra. Dominant periods in the DH-11 record are indicated in kyr; in the ocean record, they are 100, 41 and 23 kyr. b, Comparison of DH-11 and radiation³⁹ spectra. Dominant periods in radiation are 40 and 23 kyr. For a and b, record pairs are significantly coherent at the 80% level in the frequency bands with shading.

Received 23 December 1992; accepted 27 April 1993.

- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973)
- Imbrie, J. *et al.* in *Milankovitch and Climate*, Part 1 (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269–305 (Reidel, Dordrecht, 1984).
- Martinson, D. G. *et al.* *Quat. Res.* **27**, 1–30 (1987)
- Imbrie, J. *et al.* *Paleoceanography* **7**, 701–738 (1992)
- Winograd, I. J. *et al.* *Science* **258**, 255–260 (1992)
- Ludwig, K. R. *et al.* *Science* **258**, 284–287 (1992)
- Benson, L. & Klieforth, H. in *Aspects of Climate Variability in the Pacific and the Western Americas* (ed. Peterson, D. H.) 41–59 (Am. geophys. Un. Geophys. Monogr. 55, Washington DC, 1989).
- Edwards, R. L., Chen, J. H. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **81**, 175–192 (1986).
- Martinson, D. G., Menke, W. & Stoffa, P. J. *geophys. Res.* **87**, 4807–4818 (1982).
- Edwards, R. L. & Gallup, C. D. *Science* **259**, 1626 (1993).
- Shackleton, N. J. *Nature* **362**, 596 (1993).
- Mix, A. C. *et al.* *Paleoceanography* **6**, 205–226 (1991)
- Mix, A. C. in *The Last Interglacial/Glacial Transition in North America* (eds Clark, P. U. & Lea, P. D.) 19–30 (Geol. Soc. Am., Spec. Pap. 270, Boulder, 1992).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Mix, A. C. & Ruddiman, W. F. *Quat. Sci. Rev.* **4**, 59–108 (1985).
- Broecker, W. S. *et al.* *Radiocarbon* **30**, 261–295 (1988).
- Shackleton, N. J. *et al.* *Nature* **335**, 708–711 (1988).
- Bard, E. *et al.* in *Geological History of the Polar Oceans: Arctic versus Antarctic* (eds Bleil, U. & Thiede, J.) 405–416 (Kluwer, Dordrecht, 1990).
- Pisias, N. G. *et al.* *Mar. Geol.* **56**, 119–136 (1984).
- Prell, W. L. *et al.* *Paleoceanography* **1**, 137–162 (1986).
- Ruddiman, W. F., Raymo, M. E., Martinson, D. G., Clement, B. M. & Backman, J. *Paleoceanography* **4**, 353–412 (1989).
- Johnson, R. G. *Quat. Res.* **17**, 135–147 (1982).
- Shackleton, N. J., Berger, A. & Peltier, W. R. *Trans. R. Soc. Edin: Earth Sci.* **81**, 251–261 (1990).
- Pisias, N. G., Mix, A. C. & Zahn, R. *Paleoceanography* **5**, 147–160 (1990).
- Hamelin, B., Bard, E., Zindler, A. & Fairbanks, R. G. *Earth planet. Sci. Lett.* **106**, 169–180 (1991).
- Chen, J. H., Curran, H. A., White, B. & Wasserburg, G. J. *Geol. Soc. Am. Bull.* **103**, 82–97 (1991).
- Weertman, J. J. *Glaciol.* **5**, 145–158 (1964).
- Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
- Pisias, N. G. & Rea, D. K. *Paleoceanography* **3**, 21–37 (1988).
- Clemens, S. C. & Prell, W. L. *J. geophys. Res.* **96**, 22683–22700 (1991).
- Maasch, K. A. & Saltzman, B. *J. geophys. Res.* **95**, 1955–1963 (1990).
- Short, D. A., Mengel, J. G., Crowley, T. J., Hyde, W. T. & North, G. R. *Quat. Res.* **35**, 157–173 (1991).
- Gallée, H. *et al.* *J. geophys. Res.* **97**, 15713–15740 (1992).
- Arthur, M. A. & Garrison, R. E. (eds) *Paleoceanography* **1**, 369–586 (1986).
- Hilgen, F. J. *Earth Planet. Sci. Lett.* **104**, 226–244 (1991).
- Baksi, A. K., Hsu, V., McWilliams, M. O. & Farrar, E. *Science* **256**, 356–357 (1992).
- Tauxe, L., Deino, A. D., Behrensmeier, A. K. & Potts, R. *Earth planet. Sci. Lett.* **109**, 561–572 (1992).
- Jenkins, G. M. & Watts, D. G. *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, 1968).
- Berger, A. & Loutre, M. F. *Quat. Sci. Rev.* **10**, 297–317 (1991).

ACKNOWLEDGEMENTS. This work was supported by the NSF. We thank S. Clemens, R. Fairbanks, J. M. Landwehr, N. Shackleton and I. J. Winograd for comments on an earlier draft, and N. G. Pisias, W. Prell, and J. R. Toggweiler for discussions. A Duffy made the calculations required to transfer the Devils Hole chronology.

Temperatures in the Earth's core from melting-point measurements of iron at high static pressures

R. Boehler

Max-Planck-Institut für Chemie, Postfach 3060, 6500 Mainz, Germany

THE temperature distribution in the Earth's core places important constraints on the Earth's internal heat budget and on models of the geodynamo. The solid inner core crystallizes from a liquid outer core, consisting mainly of iron alloyed with a lighter element, at a depth of about 5,100 km (corresponding to a pressure of about 3.3 Mbar). Thus, the most reliable means of determining the temperature gradient in the core is to estimate the melting temperature of iron and iron-rich compounds at the pressure of the inner core boundary. Current estimates range from about 4,000 to 8,000 K; but these estimates, obtained from shock compression¹⁻³, theory (discussed in ref. 4) and extrapolation of static pressure data^{2,3,5}, are poorly constrained. Here I present melting-point measurements on iron and iron-oxygen compounds at static pressures of up to 2 Mbar. Extrapolation of these results to 3.3 Mbar yields a temperature at the inner-core boundary of $4,850 \pm 200$ K. A weak change in optical absorption observed above 2,000 K may correspond to the solid-solid phase transition found in shock experiments at 2 Mbar (ref. 1).

It has been shown⁵ that previous estimates of the melting point of iron from both static and shock experiments^{2,3} are too high, and that the phase diagram of iron is far more complicated than previously assumed. These newer measurements at pressure up to 1.2 Mbar using improved techniques have since been reproduced at lower pressures⁶⁻⁹ and suggest a triple point between the hexagonal close-packed (h.c.p., ϵ), face-centred cubic (f.c.c., γ) and liquid phases near 1 Mbar and 3,000 K. But these results require a very large melting gradient above this triple point if extrapolations are to agree even with the lowest temperatures estimated for shock melting at 2.4 Mbar (ref. 1). An additional consequence of this triple point is that the solid-solid phase transition found in those shock data at 2 Mbar must be due to a previously unknown phase^{10,11}.

Here I extend the phase diagram of iron to 2 Mbar, into the pressure regime of shock-melting experiments, nearly doubling the pressure range of previous static melting experiments. The experimental technique is essentially the same as that described in detail previously^{5,12}. A cross-section of the high-pressure cell assemblage is shown in Fig. 1. Higher pressures were obtained by technical improvements a Nd-yttrium-lithium-fluoride laser (Coherent, Antares-YLF) with better beam pointing, power stability and mode quality than the Nd-YAG laser used previously allowed more reliable *in situ* visual detection of the solid-solid transitions and more reproducible melting measurements. At constant laser power, the largest observed temperature fluctuation at about 3,000 K was ± 10 K for exposure times of 0.01. In addition, a more sensitive CCD-spectrometer (Chromex ISP250 / SI:TE/CCD-576E) with improved signal-to-noise ratios allowed faster temperature measurements from small areas.

Temperatures were measured from areas of $1 \mu\text{m}$ diameter in the spectral range 550–800 nm (576 diodes) with an exposure time of 0.1 s. Planck's radiation function was fitted to the spectra and solved for both temperature and emissivity assuming wavelength-independent emissivities. At high pressure, the relationships between emissivity and wavelength is unknown, but

for metals this correction at 1 atm is nearly independent of temperature and is typically less than 100 K for the present spectral range. Checks on wavelength-dependent emission were made by fitting different portions of the emission spectrum at constant temperature over a large spectral range, and yielded no significant deviations.

Melting was determined visually as the onset of convective motion during increasing laser power, while temperatures were continuously monitored. Melting temperatures thus measured agree with other methods such as changes in reflectivity, and discontinuities in the laser power-temperature function⁵. Repeated heating of the same portion of the sample yielded no systematic changes in the melting temperatures, so effects of chemical reactions between molten iron and sapphire are probably minimal. The melting temperature ranges shown in Fig. 2 represent the variation in the observed onset of melting for at least five different experiments at a single pressure. The surface of the sample was frequently examined under a microscope with a $350\times$ magnification. After melting, round, shiny melting textures were identified. Discoloration, an indication of oxidation of iron, was only observed if the sample was not carefully dried.

The present melting measurements cover a pressure range from 0.54 to 1.97 Mbar, and agree with previous measurements^{2,3} to 1.15 Mbar. Above 1 Mbar and 2,900 K, the melting slope increases from ~ 0.3 to 1.0 K kbar⁻¹, a strong indication of a triple point which was predicted previously^{10,13,14} at that pressure. To test this, I extended measurements of the h.c.p.-f.c.c. (ϵ - γ) transition to over 700 kbar. The slope of this transition is well constrained by resistivity measurements to over 400 kbar (see Fig. 2a). The high laser stability allowed *in situ* detection of this phase change by visual observation of small but reproducible changes in the optical absorption, and by changes in the slope of heating power against temperature. Extrapolation of the new data, which agree with the previous resistivity experiments, results in an intercept with the melting curve at about 1 Mbar.

The thermodynamically calculated temperature in the shock-melting experiment of Brown and McQueen¹ at 2.4 Mbar is $5,500 \pm 500$ K. Linear extrapolation of my melting data from 2 to 2.4 Mbar yields $4,250 \pm 300$ K, including a maximum systematic error of ± 100 K due to the unknown emissivity-wavelength dependence.

The melting curve contrasts with earlier work by Williams *et al.*^{2,3} who combined laser-heating data to 1.1 Mbar and direct shock temperature measurements to 2.5 Mbar. Systematic

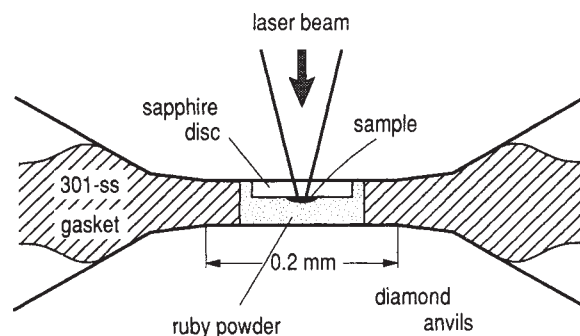


FIG. 1 Schematic cross-section through a diamond cell assemblage. The sample is in direct contact with a polished disc of sapphire, reducing scattering of light, pressure drops on heating, and intergranular migration of molten iron. Before applying pressure, the high pressure cell was dried at 120°C for several hours in a vacuum furnace and subsequently flushed with dry argon. Pressures were measured using the ruby pressure scale¹⁸ by sampling from areas $1 \mu\text{m}$ in diameter in the ruby powder matrix surrounding the sample before and after each heating experiment in the immediate vicinity of the hotspot. No correction was made for thermal pressures.

errors in these experiments, however, were large. Those in the static experiment are discussed in detail in ref. 5. Shock temperature measurements are subject to large uncertainties because of the unknown thermal conductivity and optical properties of the window material during the shock. Furthermore, the short rise times for pressure and temperature may cause the melting curve to overshoot, and equilibrium conditions may be offset.

Brown and McQueen¹ found evidence of a solid-solid transition in shocked iron at 2 Mbar and $4,400 \pm 300$ K, which

they interpreted as the ϵ - λ transition. From our present knowledge of the phase diagram of iron, however, this transition must be due to a new, yet unknown phase. I observed a weak sign of a change in optical absorption during temperature cycling above about 2,000 K (Fig. 2). This face boundary from ϵ to the unknown face is shown in Fig. 2 up to 1.6 Mbar.

The chemistry of the Earth's core is still under dispute. To estimate the upper bounds of the outer core liquidus, one must consider the depression of the melting temperature of iron due to light elements such as oxygen, sulphur, hydrogen and carbon. For initial measurements at core pressures, I chose mixtures of iron and iron oxide, likely candidates to dominate outer core chemistry¹⁵. Samples with compositions Fe:8 wt% FeO and Fe:30 wt% FeO with grain sizes of a few micrometres were synthesized in a piston-cylinder apparatus at 10 kbar and 800 °C (B. Hibberson and A. E. Ringwood). Melting experiments on Fe_{0.92}FeO_{0.08} done in an argon pressure medium to 581 kbar, and on Fe_{0.07}FeO_{0.03} to 260 kbar in argon and to 1.4 Mbar in the pressure medium described in Fig. 1. At pressures below about 500 kbar, solid particles could be clearly observed in a molten matrix at the onset of melting, whereas inhomogeneity could not be observed at higher pressures. The results at low pressures shown in Fig. 2b show that the mixtures have lower melting temperatures than pure iron and FeO, whereas above about 600 kbar, melting depression disappears within the uncertainty of the measurement. This could be due to the inability to detect the solidus temperature at pressures above about 600 kbar for strongly pressure-dependent eutectic compositions, or to a change from a eutectic to a solid-solution system. These uncertainties are reflected by the relatively large error bars, but it seems certain that melting depression plays a minor role in the Fe-FeO system at very high pressures. Ringwood and Hibberson⁶ estimated the eutectic composition and temperature of this system at 160 kbar at 10 wt% FeO and $1,943 \pm 20$ K, respectively. My melting data agree with these results at the same pressures. My melting data agree with these results at the same pressures.

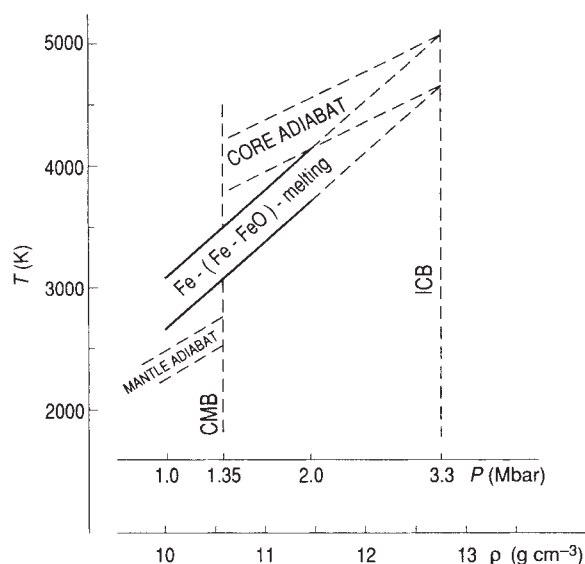
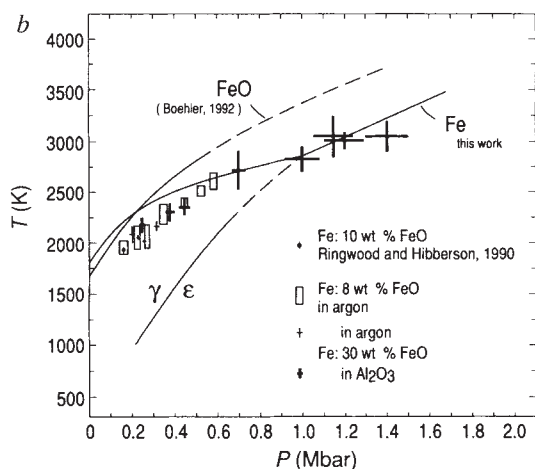
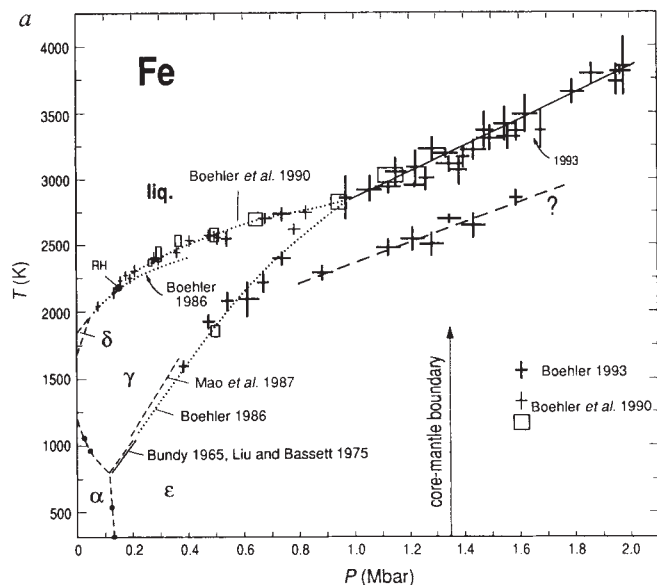


FIG. 2a, Phase diagram of iron to 2 Mbar. The low-pressure phase diagram is taken from Boehler *et al.*^{2,3} and references therein, for the point labelled RH (ref. 6). Bold symbols represent new measurements. The data marked with a question mark represent weak changes in the optical absorption of iron during temperature cycling. Temperature ranges represent the variation in the measured onset of melting for at least five temperature cycles. b, Melting data in the system Fe-FeO shown with the previously measured melting curve of FeO (ref. 11) and the present iron phase diagram. At pressures below about 600 kbar the data represent the onset of melting of a matrix including solid particles. The larger temperature uncertainties at higher pressure may be due to the solidus-liquidus temperature range.

FIG. 3 Melting and adiabatic gradients in the outer core of the Earth. The mantle temperature at the core-mantle boundary (CMB) is estimated as 2650 ± 100 K (ref. 11). The present melting data between 1 and 2 Mbar are linearly extrapolated against density to the inner core boundary (ICB). The core temperature at the CMB is calculated from an outer core adiabat using $(\partial T / \partial P)_S = \alpha T / C_p \rho$. This yields ΔT across the CMB of 1,350 K.

Extrapolation of the melting temperatures to the inner-core boundary (ICB) at 3.3 Mbar can only yield an approximate temperature for the Earth's inner core because of the simplifications required in the absence of better knowledge. The following assumptions were made. (1) The relationship between density¹⁶ is linear because this relationship best describes previous melting experiments of alkali metals, of iron between triple points and of iron oxides¹²; and because the compression range of iron between 2 and 3.3 Mbar is only 11%. (2) There are no triple points between 1 and 3.3 Mbar in the Fe-FeO system. (3) The melting slope of pure iron above 1 Mbar is not significantly changed by the presence of other elements. (4) Eutectic melting depression plays a minor role for the outer core. Disregarding possible large uncertainties in these assumptions, this extrapolation yields a temperature of $4,850 \pm 200$ K at the ICB (Fig. 3) compared to $5,800 \pm 500$ K (ref. 1) and $7,800 \pm 500$ K (refs 2,3) based on shock measurements.

The temperature jump across the core-mantle boundary plays a key role in mantle dynamics. It can be calculated from my estimate of the ICB temperature and the adiabatic gradient across the outer core from

$$(\partial T / \partial P)_s = \alpha T / C_p \rho$$

using average values for outer-core conditions of thermal expansivity $\alpha = 5 \times 10^{-6} \text{ K}^{-1}$ (ref. 5) temperature $T = 4,400 \pm 200$ K, specific heat $C_p = 3R$ (a lower limit) and density⁴. The value of $(T/P)_s$ does not significantly change across the liquid-solid

transition for the alkali metals at high compression¹⁷, and the volume dependence of the adiabatic gradient is the same for a large class of materials¹⁸. Therefore, the value of the adiabatic temperature drop across the outer core is calculated for solid iron, and yields 850 K. From this, the current best estimate of the temperature at the core side of the core-mantle boundary is $4,000 \pm 200$ K compared with a mantle boundary temperature of 2650 ± 100 K (ref. 12), yielding a thermal boundary layer in excess of 1,300 K. □

Received 17 March; accepted 4 May 1993.

1. Brown, J. M. & McQueen, R. G. *J. geophys. Res.* **91**, 7485–7494, (1986).
2. Williams, Q., Jeanloz, R., Bass, J., Svendsen, B. & Ahrens, T. J. *Science* **236**, 181–182 (1987).
3. Williams, Q., Knittle, E. & Jeanloz, R. *J. geophys. Res.* **96**, 2171–2184 (1986).
4. Anderson, O. L. *Geophys. J. R. astr. Soc.* (1986).
5. Boehler, R., von Bargen, N. & Chopelas, A. *J. geophys. Res.* **95**, 21731–21736 (1990).
6. Ringwood, A. E. & Hibberson, W. *Phys. Chem. Miner.* **17**, 313–319 (1990).
7. Shen, G., Lazor, P. & Saxena, S. K. *Phys. Chem. Miner.* (in the press).
8. Lazor, P., Shen, G. & Saxena, S. K. *Science* (in the press).
9. Yoo, C. S., Akella, J. & Ruddle, C. M. *Eos* **64** (1992).
10. Boehler, R. *Geophys. Res. Lett.* **13**, 1153–1156 (1986).
11. Ross, M., Young, D. A. & Grover, R. J. *geophys. Res.* **95**, 21713–21716 (1990).
12. Boehler, R. *Earth planet. Sci. Lett.* **111**, 217–227 (1992).
13. Birch, R. *Geophys. J. R. astr. Soc.* **29**, 373–387 (1972).
14. Lui, L. *Geophys. J. R. astr. Soc.* **43**, 697–705 (1975).
15. Ohtani, E., Ringwood, A. E. & Hibberson, W. *Earth planet. Sci. Lett.* **71**, 94–103 (1984).
16. Kraut, E. A. & Kennedy, G. C. *Phys. Rev. Lett.* **16**, 608–609 (1966).
17. Boehler, R. *Phys. Rev.* **27**, 6754–6762, (1983).
18. Chopelas, A. & Boehler, R. *Geophys. Res. Lett.* **19**, 1983–1986 (1992).
19. Mao, H. K., Bell, P. M., Shaner, J. W. & Steinberg, D. J. *J. appl. Phys.* **49**, 3276–3283 (1978).

ACKNOWLEDGEMENTS. I thank E. Ringwood and W. Hibberson for providing samples, A. Chopelas for developing software, and D. Yuen for discussions. I also appreciate comments by O. Anderson, M. Brown and M. Ross.

Amplification and sequencing of DNA from a 120–135-million-year-old weevil

Raúl J. Cano*, Hendrik N. Poinar*, Norman J. Pieniazek†, Aftim Acra‡ & George O. Poinar Jr§

* Biological Sciences Department, California Polytechnic State University, San Luis Obispo, California 93407, USA

† Parasitic Diseases Branch, Centers for Disease Control, Atlanta, Georgia 30333, USA

‡ Department of Environmental Health, American University of Beirut, Beirut, Lebanon

§ Department of Entomological Sciences, University of California, Berkeley, California 94720, USA

DNA has been successfully isolated from both fossilized plant¹ and animal tissues^{2–6}. The oldest material, dated as 25–40 million years old (Tertiary), was obtained from amber-entombed bees^{4,5} and termites⁶. Tissues from both these insects yielded DNA of good quality, which could be amplified by the polymerase chain reaction (PCR) and subsequently sequenced, including the genes encoding 18S ribosomal RNA^{5,6} and 16S rRNA⁶. We report here the extraction of DNA from a 120–135-million-year-old weevil (Nemonychidae, Coleoptera) found in Lebanese amber, PCR amplification of segments of the 18S rRNA gene and the internal transcribed spacer, and the corresponding nucleotide sequences of their 315- and 226-base-pair fragments, respectively. These sequences were used for preliminary phylogenetic analysis of the nemomychid's sequence with three extant coleopterans: *Lecontellus pinicola* (Nemonychidae), *Hypera brunneipennis* (Curculionidae) and the mealworm *Tenebrio molitor* (Tenebrionidae), and two

extant dipterans: the fruitfly *Drosophila melanogaster* (Drosophilidae) and mosquito *Aedes albopictus* (Culicidae) for the purpose of ascertaining the origin of the extracted and amplified DNA. The results revealed that the PCR-amplified material is that of the extinct nemomychid weevil. This represents the oldest fossil DNA ever extracted and sequenced, extending by 80 million years the age of any previously reported DNA^{4–6}.

We report the successful isolation and nucleotide sequencing of segments of ribosomal DNA from a 120–135-Myr-old nemomychid weevil. The fossil weevil (Fig. 1) was obtained from a piece of Middle Eastern amber⁷ from beds located near the towns of Jezzine and Dar al-Baidha in Lebanon. These beds occur in primary deposits of the Neocomian epoch of the Early Cretaceous, as well as in secondary deposits in the Neocomian epoch and the Aptian stage. These deposits are dated from 120 to 135 Myr^{8,9} and represent the oldest known amber containing insect inclusions.

The weevil has been identified as belonging to the extinct subfamily Eobelinae of the family Nemonychidae in the superfamily Curculionoidea (Coleoptera), and is being described as a new genus and species¹⁰. Nemonychid weevils are the probable sister group of all other Curculionoidea (weevils) and represent an ancient group of conifer-feeding beetles^{11,12}. Of the 22 recognized genera in the Nemonychidae, all but two develop on pollen in the male cones of conifers¹³. The greatest variety of forms occurs on species of *Araucaria* and *Agathis* in the Araucariaceae and these araucarian nemomychids appear to represent the most primitive of all extant forms and most closely resemble the fossil nemomychids of the upper Jurassic Kara Tau deposits in central Asia^{13,14}. It was concluded from infrared and mass spectrometry, as well as thin-layer chromatography¹⁵, that Lebanese amber was formed from araucarian trees. Thus, it is quite possible that the fossil weevil developed in the male cones of the resin-producing tree. Only a few beetles have been repor-

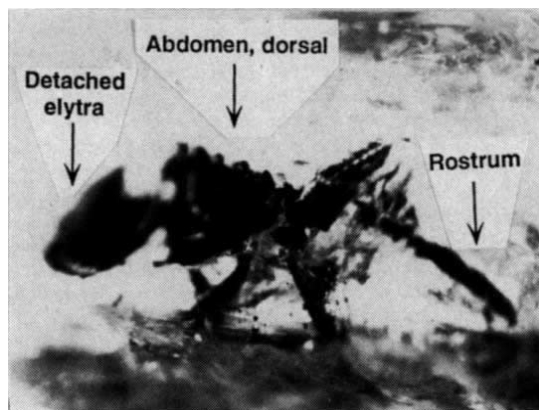


FIG. 1 Lateral view of the extinct weevil (Nemonychidae: Coleoptera) in 120–135-million-year-old Lebanese amber.

ted from Lebanese amber^{11,16} and none of them are nemonychids. In fact, nemonychids are rare in amber; none have been described and a passing mention of their occurrence in Baltic amber is the only record of this group in fossilized resin¹⁷. Today, the extant nemonychids form an amphipolar group, represented in the north and south temperate zones by disjunct groups and poorly represented in the tropics^{11–13}.

The amber piece was surface-sterilized and cracked as described⁴. During this process the amber broke open and a pellet of tissue was removed from the weevil's body cavity. DNA was extracted using Chelex-100 (BioRad), which has been used successfully to extract high-quality DNA from tissues and forensic materials for enzymatic amplification^{18,19}. DNA was also extracted from alcohol-preserved tissues of the weevil *H. brun-*

FIG. 2 Sequence alignment for a 315-base-pair sequence resulting from the 5' region of the 18S rRNA gene (a) and a 226-base-pair sequence from the internal transcribed spacer (ITS) of the rRNA gene (b) of 120–135-million-year-old weevil srRNA. Dots indicate that the base is identical to that of the outgroup *Artemia salina*. Base substitutions are indicated by the first letter of the substituting base, and gaps are indicated by a dash. Nucleotide sequences from the 18S rRNA gene used in phylogenetic inference studies are indicated by an asterisk. Sequences from the ITS region are shown with the fossil nemonychid on top to illustrate the sequence similarities between the extinct and extant weevils.

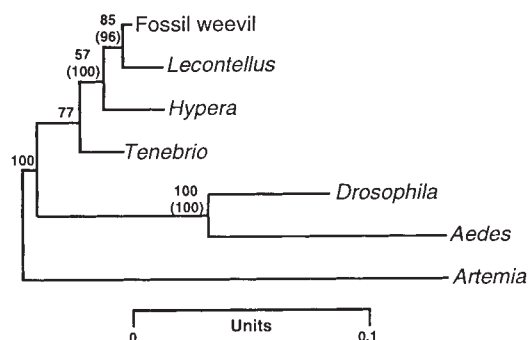
METHODS. DNA was extracted using Chelex 100 (BioRad, Richmond, CA). The pellet of tissue from the body cavity of nemonychid weevil measuring 1–2 mm in diameter and thoracic tissue from extant weevils were suspended in 500 μ l of a 5% aqueous suspension and incubated for 2–4 h at 56 °C with continuous shaking. Samples were then mixed for 30 s and placed in a dry bath at 95 °C for 5 min. After centrifugation at top speed in a microcentrifuge for 5 min, the supernatant containing extracted DNA was stored at –20 °C in a sterile 0.7- μ l microcentrifuge tube. For PCR amplification, extracted DNA was diluted 1:10 and 1:20 in filter-sterilized double-distilled H₂O. Aliquots (5 μ l) of the diluted sample were used as the source of DNA. Symmetric PCR amplifications (1 unit of *Taq* DNA polymerase (Boehringer-Mannheim), 2 μ g ml^{–1} bovine serum albumin, 0.5 μ M each of forward and reverse primers, 2.0 mM MgCl₂ and 0.2 mM dNTPs in 50 μ l) were reacted in a programmable thermal cycler (Ericomp, La Jolla): DNA templates were denatured for 2 min at 95 °C, followed by a primer annealing step at 48 °C for 30 s, and an extension step of 60 s at 72 °C. The following 30 cycles consisted of a 30-s denaturation step at 94 °C, a 30-s primer annealing step at 48 °C, and a 60-s primer extension step at 72 °C. The last cycle consisted of a 30-s denaturation step, a 30-s primer annealing step, and a 5-min primer extension step. The products of each amplification were evaluated by 1.2% agarose or 5% polyacrylamide gel electrophoresis. Each amplification reaction mixture (25 μ l) was purified using a Magic PCR purification kit (Promega) to remove excess dNTPs, primer and buffer components. 10 μ l purified PCR product was used as template in an asymmetric PCR to generate a single-stranded DNA template for sequencing. Amplification conditions were identical to those described for symmetric PCR, except that one of the primers was used at 0.5 μ M and the other at 0.02 μ M. The low-concentration primer was also used in the sequencing reaction. Single-stranded DNA products were sequenced using a Sequenase II DNA sequencing kit (USB, Ohio) with [α -³⁵S]dATP according to the manufacturer's instructions. Electrophoresis of sequencing products were in a 6% sequencing gel (Gel-Mix 6, Gibco). Autoradiographs of air-dried sequencing gels were made using XAR X-ray film (Kodak). Sequences were inferred from autoradiographs using a gel reader (IBI, New Haven CT).

a

Artemia	TTATGTTCC	TTAGATCGTA	CTATATCCTA	CTTGGAATAC	TGTGGTAAT	CTAGAGCTAA
Drosophila	T	T	T	T	T	T
Aedes	C...AA...TA	CA...A...T	TA...CTAGT	T	T	T
Tenebrio	T	T	T	T	T	T
Hypera	T	T	T	T	T	T
Lecontellus	T	T	T	T	T	T
Nemonychid	T	T	T	T	T	T
Artemia	TACATGCACA	-ATAGCCCCA	ACTTC--ACG	GAAGGGTGGC	TTTATTAGA	TCAGAGCAAA
Drosophila	AT	T.A.A.ATG	..C.T...T	..G.C.T...	G	CT...A...
Aedes	A-	..AT..AGG	..C...G...	..CCT...	AA.....T	C--A....
Tenebrio	A	..C.G...T	T...	A.C...	T	T
Hypera	A	..C.G...T	T...	A.C...	T	T
Lecontellus	A	..C.G...T	T...	A.C...	T	T
Nemonychid	A	..C.G...T	T...	A.C...	T	T
Artemia	TCGG-GGC--	TTGGGCTCGT	CT-----	CTTGGTGAAT	CTGAATAACT	ATA-GCCGAT
Drosophila	G...ATC...AA	G-----A	..GTTATATTG	T	T	T
Aedes	...TCCT..CG	A	..GCTGGA...	G...A--A...	..G...T...	...T...T...
Tenebrio	...T...GG	..CTC--	..ATCGTACAA	T	T	...T...T...
Hypera	...T...GG	G...TA...	..ATCGTACAA	T	T	...T...T...
Lecontellus	...T...GGG	G...TA...	..ATCGTACAA	T	T	...T...T...
Nemonychid	...T...GG	G...TA...	..ATCGTACAA	T	T	...T...T...
Artemia	CGCAGCGTCT	CGCAGCGGCG	ACGTGCTCTT	CAATGCTCTG	CCTTATCAAC	TTTCGATGGT
Drosophila	..T.T....	T.T....A	..AGA...	A...	..C.....	..T.....
Aedes	..T.T....	T.T....A	..AGA...	A...	..C.....	..T.....
Tenebrio	T	T	T	T	T	T
Hypera	T	T	T	T	T	T
Lecontellus	T	T	T	T	T	T
Nemonychid	T	T	T	T	T	T
Artemia	AGGCTATGCG	ACGGGTACAC	GGGAATCGGG	GTTCGATTCG	GGAGAGGAG	CTTGAGAAAC
Drosophila	..TA.C.AG	..TACC.TG	TT.C.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Aedes	..TA..GAG	..TACC.TG	TT.C.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Tenebrio	..T.C....	C.TACC.TG	TT.T.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Hypera	..T.C....	C.TACC.TG	TT.T.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Lecontellus	..T.C....	C.TACC.TG	TT.T.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Nemonychid	..T.C....	C.TACC.TG	TT.T.A...	TAA..GGGAA	TC..G.TTC	AT.CC.G.GA
Artemia	GGGAGCCTGA	GAAC	T	T	T	T
Drosophila	T	T	T	T	T	T
Aedes	T	T	T	T	T	T
Tenebrio	T	T	T	T	T	T
Hypera	T	T	T	T	T	T
Lecontellus	T	T	T	T	T	T
Nemonychid	T	T	T	T	T	T

b

Nemonychid	CGTCTCCGTA	GTGAACCTGC	GGAGGGATCA	TTTATTCGGT	TGCATCAAC	AGGCAATCTT
Lecontellus	...ATC.T	..C.T..T	..A.T.C..	..G.A..C	AA.....T	GG
Hypera	...G.T.-C	..TCTT.AT	..AT.CA.GAC	CC..G..--A	..C...-G.A	..A...TC.A
Drosophila	TTG.GTG..G	AACTTG.G.A	A.GATC-ATT	A...GTATA	A-T...CTTA	CC..TTAA.A
Aedes	G.T...A	A	ATT...TT	..A.A..A.G	..AT..AT	CC...--GA
Nemonychid	CGGATTGTCT	GCAACCACTG	ACATTGTGGA	GGGGTCAACA	TTGGAATGGT	GTATTTTTAA
Lecontellus	T	T	T	T	T	T
Hypera	AAT..GTGT	AT...ACA	G	..A.T..GG	G	G
Drosophila	GA..G..G	..T.T.A	..G..G	..A.T..GG	G	G
Aedes	T	T	T	T	T	T
Nemonychid	TTTAAATCTA	TCTTTACATG	TGAGACAA	TTTGAATTA	ATCTTCAAAA	CTTTCACAAA
Lecontellus	...G...G	..C...T	T	T	T	..C.A.G.C
Hypera	...T.G...	..G.C...T	..T.A..T	..C.G.C..C	..AG..AT	..TGC.TTG.G
Drosophila	...CT...A	..AAAA..A	ATT...TT	..ACCT.T.C	CA-AA..C..	..CC.AGG..G
Aedes	...CT...A	..AAAA..A	ATT...TT	..A.A..A.G	..AT..AT	..C.A.G.CC
Nemonychid	CGGATCTCTT	GGTCTCGCA	TCGATGAAGA	ACGCAGCAAA	CTGGGT	T
Lecontellus	T	T	T	T	T	T
Hypera	T...GA..A	..G...T	T	T	T	T
Drosophila	...G...A	..C..GT.G	T	T	T	T
Aedes	T...A..A	..C..AT.GG	T	T	T	T



neipennis and dried museum specimens of the extant nemomychid *L. pinicola* in order to provide amplifiable DNA for comparative studies. No DNA was extracted from extant species until after the DNA from the fossil nemochid had been extracted, amplified and sequenced.

Extracted DNA was used as a template using a panel of primers whose sequences were derived from conserved regions of eukaryotic 18S rRNA corresponding to the 5' end of the 18S rRNA gene (NS1+NS2) (ref. 20) and the internal transcribed spacer (ITS1+ITS2)²¹. Amplified DNA of the expected size (see ref. 21) was obtained from the fossil DNA with each primer pair used. As bacterial DNA either does not amplify with eukaryotic primers or gives fragments of a different size, our results support the hypothesis that the amplified DNA was of eukaryotic origin.

Single-stranded templates from both these amplified ribosomal gene products were generated by asymmetric PCR²² and sequenced using the dideoxy termination method. For comparative analysis, sequences consisting of 315 base pairs from the NS1+NS2 amplified product and of 226 base pairs from the ITS1+ITS2 were chosen to evaluate the nemomychid origin of the extracted and amplified DNA.

These sequences were used to search the GenBank database using the IBI/Pustell DNA analysis program. The highest similarity scores were obtained with all insect sequences present in the database: the coleopteran *Tenebrio molitor* (accession number, X07801) and two dipterans: *Drosophila melanogaster* (accession numbers, M21017 and M29800) and *Aedes albopictus* (accession number X57172). An even higher score was obtained when the nemomychid weevil sequence was compared with those obtained from the extant weevils *L. pinicola* and *H. brunneipennis*. Significantly lower scores were seen with human and bacterial small RNA, demonstrating that the amplified DNA was of insect origin, and probably from curculionid insect origin. Alignment of the nemomychid weevil small rRNA sequences with the five insect sequences already mentioned is shown in Fig. 2.

FIG. 3 Phylogenetic tree, obtained by Neighbor Joining, of extinct and extant taxa based on sequences of the 18S rRNA gene showing bootstrap confidence values. Trees were obtained by Neighbor Joining (NJBOOT2 and Treeview), Maximum Likelihood (DNAML), and Maximum Parsimony (PAUP 3.0s). All three trees had identical topology. The numbers at the nodes of branches represent bootstrap confidence values obtained from 2,000 replications²⁶. Numbers in parentheses are the bootstrap confidence values obtained from Neighbor Joining analysis when ITS sequences in Fig. 2b were appended to the 18S rRNA sequences (Fig. 2a). Corrections for transitions and transversions were made using the Kimura 2 parameter distance method²⁷. Maximum Likelihood tree was constructed with the DNAML program of PHYLIP 3.4 (ref. 23) using the brine shrimp *Artemia salina* as the outgroup, with randomized data input (J option), and global rearrangement of data (G option). A total of 6 independent runs were evaluated. All resulting trees were identical. Length of Maximum Parsimony tree by Branch and Bound method was 125. Consistency index = 0.968, or excluding uninformative sites, 0.915.

The aligned sequences were prepared for phylogenetic analysis by trimming gaps and regions around them to leave informative sites. Next, phylogenetic analysis was done using Maximum Likelihood (DNAML) in the PHYLIP 3.4 package²³, Maximum Parsimony using PAUP 3.0s (ref. 24), and Neighbor Joining (ref. 25) using the programs NJBOOT2 and Treeview (K. Tamura, personal communication). Statistical confidence in the topologies was assessed using the bootstrap method, with 2,000 replications²⁶. *Artemia salina* sequence (accession number X01723). The topology of the trees obtained with all program runs was essentially identical (Fig. 3).

From this simple phylogenetic tree it can be seen that coleopterans and dipterans form distinct branches. The phylogenetic inference programs consistently placed the extinct nemomychid weevil in a sister group with the extant nemomychid weevil (*Lecontellus*) and closely related to the other coleopterans. The bootstrap values for the sister group consisting of the nemomychid weevils were notably high (85–96%). Additionally, the branch length of the fossil nemomychid weevil is shorter (closer to the node) than the extant nemochid weevil, as would be expected for an extinct organism. These data are further evidence that the extracted DNA is from the extinct weevil rather than from contamination. Although this tree could have arisen from the relatively small number of informative sites used by us in the phylogenetic analysis, it still reflects the similarities between the nucleotide sequences of the DNA isolated from the extinct weevil with those of extant related genera. On the basis of these results, it can be inferred that the 315- and 226-base-pair fragments of small rDNA reported here originate from DNA extracted from the extinct weevil's tissue and not from contaminant DNA.

These results demonstrate the successful extraction and amplification of DNA from a 120–135-Myr-old amber-fossilized insect. We speculate that on the basis of our investigation the majority of animal remains in amber have preserved DNA that can be extracted and studied, thus making amber a treasure chest for molecular palaeontologists. □

Received 29 September 1992; accepted 30 March 1993.

- Golenberg, E. M. et al. *Nature* **344**, 656–658 (1990).
- Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A. & Wilson, A. C. *Nature* **312**, 282–284 (1984).
- Thomas, R. H., Schaffner, W., Wilson, A. C. & Pääbo, S. *Nature* **340**, 465–467 (1989).
- Cano, R. J., Poinar, H. N. & Poinar, G. O. *Jr Med. Sci. Res.* **20**, 249–251 (1992).
- Cano, R. J., Poinar, H. N. & Poinar, G. O. *Jr Med. Sci. Res.* **20**, 619–623 (1992).
- DeSalle, R., Gatesy, J., Wheeler, W. & Grimaldi, D. *Science* **257**, 1860–1862 (1992).
- Poinar, J. G. O. *Jr Life in Amber* (Stanford Univ. Press, Stanford, 1992).
- Schlee, D. & Dietrich, H.-G. *Neues Jahrb. Geol. Paläont. Monatshefte* (Stuttgart) **1**, 40–50 (1970).
- Schlee, W. & Glöckner, W. *Bernstein* 1–72 (Stuttgarter Beitr. Naturk., Germany, 1978).
- Kuschel, G. & Poinar, G. O. *Jr A. Ent. Scand.* (in the press).
- Crowson, R. A. *The Biology of Coleoptera* (Academic, London, 1981).
- Crowson, R. A. *Entomol. Gener.* **11**, 57–58 (1985).
- Kuschel, G. *Geogr. J.* **7**, 499–504 (1983).
- Arnoldi, L. V. *Trudy Paleont. Inst.* **161**, 142–176 (1977).
- Bandel, K. & Vavra, N. N. *Jb. Geol. Paleont. Mh.* **1**, 19–33 (1981).
- Whalley, P. *Insects from Lebanese Amber* (British Museum of Natural History (Geol.), London, 1981).

- Crowson, R. A. *Atti X Congr. Nazionale Italiano di Entomologia Sassari* 1–90 (1974).
- Walsh, P. S., Metzger, D. A. & Higuchi, R. *Biotechniques* **10**, 506–513 (1991).
- Jung, J. M., Comey, C. T., Baer, D. B. & Budowle, B. *Int. J. Legal Med.* **104**, 145–148 (1991).
- Dams, E. et al. *Nucleic Acids Res.* **16** (suppl.), r87–r173 (1988).
- White, T. J., Bruns, T., Lee, S. & Taylor, J. in *PCR Protocols, a Guide to Methods and Applications* (ed. Innes, M., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 315–324 (Academic, New York, 1990).
- Gyllenstein, U. B. & Erlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7652–7656 (1988).
- Felsenstein, J. *Cladistics* **5**, 164–166 (1989).
- Swofford, D. *Illinois Nat. Hist. Survey, Champaign, Illinois* (1991).
- Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).
- Hedges, S. B. *Molec. Biol. Evol.* **9**, 366–369 (1992).
- Kimura, M. *J. molec. Evol.* **16**, 111–120 (1980).

ACKNOWLEDGEMENTS. We thank J. W. Taylor for advice and for the srRNA primers; S. Blair Hedges for advice on phylogenetic analysis; L. Etzell and J. Doyen for samples of extant weevils; H. Hardcastle for use of facilities; and D. Pieniazek for critically reviewing the manuscript.

Honest signalling among gametes

Mark Pagel

School of Mathematical Sciences, Queen Mary and Westfield College,
University of London, London E1 4NS, UK

THE gametes of many lower eukaryotic organisms emit pheromones that attract gametes of the opposite mating type or sex¹⁻⁴. Gametes move or grow in the direction of the highest pheromone concentration, suggesting that the strength of the pheromonal signal is used to infer proximity, or that the strongest signal is most likely to be noticed. Here I offer a new explanation of pheromonal signalling and chemotaxis in gametes. I show that pheromonal signals can be interpreted as sexually selected traits that honestly advertise variation in quality among gametes, given that signals are costly to produce and that gametes compete; by 'quality' I refer to some aspect of a gamete's fitness. A gamete's preference for a mating partner, then, is predicted to vary with the quality of a prospective partner as inferred from the strength of its signal. This view can explain characteristics of the signalling and mate selection behaviours of gametes that are not predicted by models of mate choice based on proximity or 'passive attraction' to the strongest signal⁵. These include repeated partner exchanges², escalated exchanges of mating pheromones⁶⁻⁹, and rejection of gametes that signal at low levels^{8,9}.

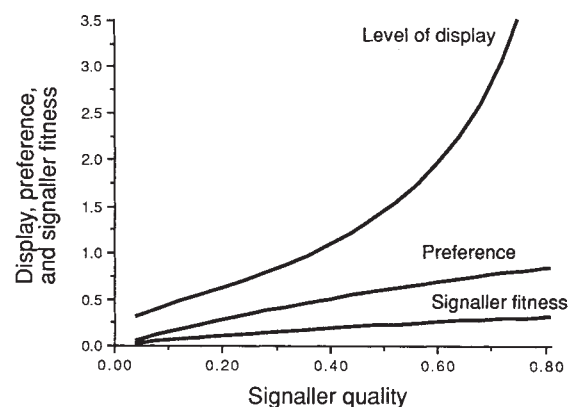
Many lower eukaryotic organisms release their gametes into a medium where to obtain a mating they must seek out the gametes of opposite mating type or sex. The isogamous gametes of the yeast *Saccharomyces cerevisiae* produce pheromones specific to their mating type, and which attract gametes of the opposite mating type^{1,7-10}. In *S. cerevisiae* growth of a gamete towards another gamete is oriented in the direction of the highest pheromone concentration, and the same pheromone induces the expression of gene products necessary for eventual fusion of gametes⁷⁻⁹. Similar systems are reported for a variety of algae,

fungi and protozoa^{1,3,4,11}. One explanation of pheromonally induced chemotaxis or chemotropism is that the gametes of these lower organisms use the strength of a signal to infer the proximity of another gamete. Alternatively, this behaviour may arise because gametes are most likely to notice and move towards the strongest signal—a process of apparent mate choice based on 'passive attraction'⁵.

An alternative view is based on the idea that systems of passive attraction, including inferring proximity, may be evolutionarily unstable. This is because they are beneficial on average to the strongest signallers, but may not result in the signal receiver acquiring the best mate. If, however, there is an association between a gamete's level of signalling and some aspect of its fitness or quality, then gamete signal receivers might benefit from actively comparing signallers and choosing the one with the strongest signal. Gametes that actively choose among mates in such circumstances would have higher fitness than non-choosy gametes, if the costs of choice are not great⁵. What form would such a signalling and choosing system take? An answer can be derived from the theory of biological signals: the strength or nature of the signal may indicate something about the quality of the signaller, if the signal is costly to produce¹². Here I use these ideas combined with a general method for the mathematical study of stable biological signalling systems¹³ to show how costly signals produced by gametes can be viewed as an honest indication of gamete quality, and that gametes of one mating type or sex can use these signals to decide how to choose among gametes of the opposite mating type or sex. The mating behaviour of gametes from lower eukaryotic organisms supports this view and is discussed below.

Consider a diploid single-celled organism such as *S. cerevisiae* or *Chlamydomonas* which produces two haploid mating types. The gametes are assumed to vary in some quality (cytoplasmic volume, for example) of interest to gametes of the opposite mating type. This quality can be genetically or environmentally determined. Gametes know their own quality, but must infer

FIG. 1 Plot of the level of display by signallers, the level of preference of receivers, and the fitness of signallers versus signaller quality. I assume that signaller fitness is given by $W_s(p, d, q) = pq^d$ where p is the receiver preference, q is the quality of the signaller, and d is the level of display (for $0 < q < 1$). Costs of display therefore increase with increasing display, but less so for higher quality individuals. Pheromonal signals often contain lipids, glycoproteins, or peptides, and so represent a cost to the gamete even if produced as a by-product of some other metabolic function. I assume that the fitness of a receiver is a function of how closely its preference tracks the true quality of the signaller, but that there is a cost to being choosy (that is, having a preference). Inferred quality is assumed to translate into a preference such that a higher level of perceived quality translates into increasing the effort put into mating with the particular signaller, or into a tendency to turn down signallers of lower perceived quality. The effort due to preference is assumed to be costly. Therefore, let receiver fitness $W_r(p, q) = p(1 - \exp(-bq))$, and the cost to having a preference by $C(p) = cp^s$, where b , c and s are arbitrary constants (I make no assumptions here about the distribution of signaller quality because I am interested in the ESS rule linking preference to quality). Increasing the value of s increases the cost of preference p . The benefits to a signaller of altering the receiver's preference are assumed to accrue more rapidly at lower levels of quality. Signallers and receivers are free to choose their level of display and their preference rule, respectively. Signallers will want to make receivers think that they are of high quality, but will face the costs of displaying. It will pay receivers to evolve a preference rule that correctly identifies variation in fitness among signallers as a function of their level of display. We want to discover the optimal rules describing how gametes should signal their quality, and how they should interpret those signals in others. A known value for the level of display as a function of quality is required to solve the differential equation (1). I assume that the lowest quality individuals cannot display at all, so that $d=0$ for $q=q_{\min}$, where $q_{\min}$ is the lowest level of quality, taken here as $q_{\min}=0.001$. Then, solving the fitness functions and substituting the values into equations (1) and (2) yields for the ESS preference and level of display: $p^* = [(1 - \exp(-bq))/sc]^{(1/s-1)}$, and $d^* = (p^*_{q=q_{\min}} - p^*)/p^*(\ln(q))$,



where the asterisk is used to denote the ESS value. The curves in the figure were drawn using $b=1$, $c=1/3$, and $s=2$. To ensure that these are local maxima, it is necessary to inspect the second-order derivatives. These results are local but not necessarily global optima. Despite placing no restrictions on the display and preference rules, the ESS level of display is determined strictly by one's quality. A low-quality gamete could signal at a high level in an attempt to masquerade as a high-quality gamete, but would suffer high costs and reduced fitness if it did so. Level of display honestly advertises quality. Level of display increases at an accelerating rate at higher and higher quality, because the costs of a given level of display are lower for higher-quality gametes. The ESS preference, in turn, is a strictly increasing function of display. The model generating these rules is a phenotypic one but requires the assumption, in the context of evolution, that the optimal rules are genetically based.

the quality of others from a signal, such as a pheromone. Signalling is assumed to be costly. A gamete's fitness as a signaller, then, will be some function of its quality, how strongly it signals, and how it is perceived by receiving gametes: let $W_s(p, d, q)$ be the fitness of a signaller of quality q , that signals at level d , and is perceived as being of quality p . A gamete's fitness as a receiver will be a function of the rule it uses to infer a signaller's quality from its level of display. Inferred quality is assumed to translate into a mating preference, and I assume that preferences are costly (in comparison to random mating). Let $W_r(p, q)$ be the fitness of a receiver that attributes quality p to a signaller of quality q , and let $C(p)$, be the costs associated with level of preference p . Gametes are free in the model to choose any level of display, and any preference rule. We seek a stable set of rules, however, such that neither the signallers nor the receivers can gain from a unilateral change in the behaviour of the other. One rule will specify how signallers should signal as a function of their quality, $d = G(q)$, and the other how receivers should perceive a given level of signalling, $p = F(d)$. Such a set of rules specifies an evolutionary stable strategy, or ESS¹⁴. These rules are unknown and must be discovered from the ESS conditions.

To understand intuitively the ESS conditions, consider a signaller that increases its level of signalling by some arbitrarily small amount. This changes receivers' preference by $p' = F'(d)$, where the prime is used to denote the simple derivative of F with respect to d . For each unit of change in p , signaller fitness changes by an amount given by $\partial W_s / \partial p$. Increasing the signal decreases the signaller's fitness, however, because of the costs of displaying. The reduction in fitness is given by $\partial W_s / \partial d$ as the increase in level of signalling becomes vanishingly small. At the ESS, the costs and benefits of a change in the level of signalling by the signaller balance each other, and thus we can define the ESS signalling condition as satisfying the equality

$$F'(d)[\partial W_s / \partial p] + \partial W_s / \partial d = 0 \quad (1)$$

The fitness of the receiver will be an increasing function of how well it can infer the quality of the signaller, but a decreasing function of the cost of preference. The ESS preference rule will balance these two effects on fitness for any given level of q . Thus, at the ESS the following equality must hold

$$\partial W_r / \partial p + C'(p) = 0 \quad (2)$$

where the second term in equation (2) describes the reduction in fitness because of costs of having a particular mate preference.

I give a solution in Fig. 1 to the differential equations in (1) and (2). At the ESS, a gamete's quality puts an upper limit on the level at which it is optimal for it to signal; signals honestly advertise signaller quality. Highest quality gametes are forced to signal at increasingly exaggerated levels to distinguish themselves from lower quality gametes. A receiver's optimal level of preference increases with the level of display: receivers can rely on the level of display to indicate quality because of the costs

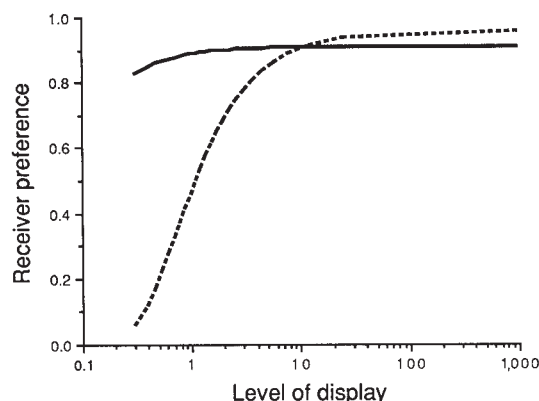
of signalling. This signalling system is evolutionary stable because any signaller displaying at a level other than the optimal for its quality, and any receiver adopting a different preference rule, will have lower fitness than gametes following the optimal rules (fitness curve in Fig. 1).

The functions plotted in Fig. 1 provide an evolutionary framework within which to interpret the mate choice behaviours of gametes. If mates are found solely from passive attraction, then we should not see evidence of active mate choice: a gamete should fuse with the first gamete of the correct species and mating type or sex that it finds. The mating types of the unicellular green alga *Chlamydomonas* are motile and are attracted to the opposite mating type^{2,15}. Upon encountering each other, however, the mating types exhibit a characteristic 'mating-type reaction', featuring multiple and repeated partner exchanges between gametes of opposite mating type². This occurs even among mixtures of gametes from only a single species, ruling out simple species recognition, and among mixtures containing many cells of both mating types² (the expected number of partner exchanges before randomly meeting one of opposite mating type is $1/P$, where P is the proportion of gametes of the opposite mating type, or 2 if the ratio of mating types is 1).

Given a choice among cells of the opposite mating type, *S. cerevisiae* gametes grow towards the one producing the strongest signal, but choose apparently at random when all cells signal at the same level^{8,9}. The mating types of *S. cerevisiae* are denoted as a and α respectively. Upon cell-cell contact, the α -gamete's pheromone induces the a -cell to produce more α -receptors¹⁶ and to secrete more of its a -pheromone in response^{9,17,18}. The induced a -pheromone is hydrophobic¹⁹ and is exchanged between the two gametes principally across their cell walls¹⁷, a characteristic that may positively identify the signaller. The induced a -pheromone then causes the α -cell to produce more α -receptors and to secrete more α -pheromone²⁰, and so on; the mating types engage in a reciprocal and escalating exchange of pheromones⁷.

Chemotropism, cell-cell contact, and the reciprocal exchange of pheromones do not ensure that the two cells will mate. Rather, if one of the mating types does not produce a sufficient induced response, it may be rejected by the other cell in favour of one that signals more strongly⁸. Gametes producing the highest levels of induced pheromone are most likely to retain a mate⁸ and induced levels up to ten times higher than the non-induced or constitutive levels have been observed¹⁷. But a gamete will fuse with another gamete that produces relatively low levels of induced pheromone if there is no other induced cell signalling at a higher level⁸. At least part of the signal released by a gamete during the reciprocal exchanges, then, may be pure 'display', that is, not strictly physiologically necessary to induce conjugation: escalated exchanges may be how gametes examine each other close up. Chemotactic agents secreted by gametes to attract gametes of the opposite type, and the reciprocal induction of pheromones between gametes have been reported for other

FIG. 2 Plot of the optimal level of preference as a function of the level of display, for two different values of the cost of preference: dashed curve corresponds to low costs, solid curve corresponds to higher costs. Increases in preference tail off for very high levels of display because the highest quality gametes are forced to signal at increasingly exaggerated levels: the gains to the receiver of preferring a high quality signaller do not vary with the signaller's level of display, but rather with its quality. The shape of the preference curve provides a possible evolutionary interpretation for patterns of receptor saturation: progressive and finally complete receptor saturation may be the mechanism by which increases in preference are measured and eventually reach an asymptote. Increasing the costs to a gamete of having a preference (increasing s in the fitness function for equation (2)), causes the ESS preference not to vary so strongly with level of display: gametes become less choosy.



fungi, and for algae and protozoa^{1,3}.

Female gametes of the anisogamous aquatic fungus *Allomyces* release a pheromone called 'sirenin' that attracts the male gametes^{11,21}. Male gametes are attracted in numbers directly proportional to the signal strength for concentrations of sirenin between 0 and 7.5 nM, above which the increase in numbers of male gametes slows²¹. The linear part of the response may reflect the number of male gametes that detect the signal, but does not explain why the rate of increase in the numbers of new male gametes attracted to the female's signal should slow down at high levels of sirenin. It can be seen from Fig. 2, however, that receiver preference should tail off at higher levels of display if the signal strength is used to infer quality.

Several factors are expected to act against the evolution of honest signalling among gametes. Gametes should not be choosy where the costs of seeking out a mate are high (Fig. 2). In the multicellular *Achlya* and *Mucor* fungi, male and female mycelia grow towards each other, sometimes over long distances, in response to a mating pheromone each secretes³. Fusion follows routinely upon contact between the two mycelia, a case of apparent passive attraction. In many anisogamous systems, competition among sperm to find and mate with eggs, such as in most higher plants and animals, will mean that sperm cannot afford to be choosy. Sperm should seek out (be passively attracted to) the first signal they detect. Nevertheless, if the eggs are not widely dispersed, competition to be noticed by males will select for females to produce strong signals. For example, male *Allomyces* gametes compete over relatively immotile female gametes. Female gametes of *Allomyces* compete to attract males, and may secrete sirenin for up to six hours¹¹. Male gametes only signal, and then weakly, if after several hours they have not found a female gamete¹¹. In gametic mating systems lacking competition among eggs, the eggs need only signal at a level that males can detect.

A special case of the model of Fig. 1 is one in which signallers merely signal at the optimal level for their condition (a phenotype-limited⁵ or 'condition-dependent strategy'^{22,23}), rather than 'choose' their level of display optimally. This signalling system also favours active mate choice if higher levels of the signal are reliably associated with fitness, and if the costs of choice are not great. Signalling of quality does not exclude other functions for the pheromones emitted by gametes, such as stimulating the expression of genes necessary for conjugation, or indicating the location of the signaller. Defects in a gamete's display rule or preference rule could have marked consequences for its fitness: sterility could, for example, be the result for a gamete that adopts a much-too-strict preference rule. The idea of gametes signalling their quality can be generalized to signalling among somatic cells. But in such cases, the relatedness of cells (they will generally be clones save for somatic mutations) may mean that the most powerful selective forces are for cooperation, and so honest signalling need not be enforced as strongly as in the case of signalling among gametes. □

Received 10 November 1992; accepted 30 March 1993.

1. Crandall, M. in *Receptors and Recognition Series A* (eds Cuatrecasas, P. & Greaves, M. F.) 45–100 (Chapman and Hall, London, 1977).
2. Wiese, L. & Wiese, W. in *Symposia of the Society of Experimental Biology, Cell–Cell Recognition* (ed. Curtis, A. S. G.) 83–104 (Cambridge Univ. Press, UK, 1978).
3. Mascarenhas, J. P. in *Receptors and Recognition, Ser. B, Taxis and Behaviour* (ed. Hazelbauer, G. L.) 169–204 (Chapman and Hall, London, 1978).
4. Miller, A. M. *Biol. Fertil.* **2**, 275–337 (1985).
5. Parker, G. A. in *Current problems in Sociobiology* 173–201 (Cambridge Univ. Press, UK, 1982).
6. Cross, F., Hartwell, L. H., Jackson, C. & Konopka, J. B. *A. Rev. Cell. Biol.* **4**, 429–457 (1988).
7. Hershkowitz, I. *Nature* **342**, 749–757 (1989).
8. Jackson, C. L. & Hartwell, L. H. *Molec. cell. Biol.* **10**, 2202–2213 (1990).
9. Jackson, C. L. & Hartwell, L. H. *Cell* **63**, 1039–1051 (1990).
10. Crandall, M. in *Symposia of the Society of Experimental Biology, Cell–Cell Recognition* (ed. Curtis, A. S. G.) 105–120 (Cambridge Univ. Press, UK, 1978).
11. Pommerville, J. *Expl Cell Res.* **109**, 43–51 (1977).
12. Zahavi, A. *J. theor. Biol.* **53**, 205–214 (1975).
13. Grafen, A. *J. theor. Biol.* **144**, 517–546 (1990).
14. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, UK, 1982).

15. Musgrave, A. & van den Ende, H. *Trends biochem. Sci.* **12**, 470–473 (1987).
16. Hartig, A., Holly, J., Saari, G. & MacKay, V. L. *Molec. cell. Biol.* **6**, 2106–2114 (1986).
17. Strazdis, J. R. & MacKay, V. L. *Nature* **305**, 543–545 (1983).
18. Achstetter, T. *Molec. cell. Biol.* **9**, 4507–4514 (1989).
19. Anderegg, R. J., Betz, R., Carr, S. A., Crabb, J. W. & Duntze, W. *J. biol. Chem.* **263**, 18236–18240 (1988).
20. Hagen, D. C. & Sprague, G. F. *J. molec. Biol.* **178**, 835–852 (1984).
21. Machlis, L. *Plant Physiol.* **52**, 527–530 (1973).
22. Andersson, M. *Evolution* **40**, 804–816 (1986).
23. Pomiankowski, A. N. *Oxford Surveys Evol. Biol.* **5**, 136–184 (1988).

ACKNOWLEDGEMENTS. This work was supported by a grant from the SERC.

Contour from motion processing occurs in primary visual cortex

Victor A. F. Lamme, Bob W. van Dijk
& Henk Spekreijse*

The Netherlands Ophthalmic Research Institute, PO Box 12141, 1100 AC, Amsterdam-Zuidoost, and The Laboratory of Medical Physics, University of Amsterdam, Meibergdreef 15, 1105 AZ, Amsterdam, The Netherlands

RELATIVE motion is one of the most salient cues for segmentation of a visual scene into separate objects. This is illustrated by the vivid contours that are perceived when random dot patterns move in different directions^{1,2}. Once motion is halted in such displays the segmentation contours disappear. This makes random dot patterns ideal for the study of contour from motion processing in isolation. Contour from motion processing obviously relies on direction-selective neurons, which are found in many visual cortical areas^{3–5}. It is, however, largely unknown at what level of processing their signals interact to serve the global process of motion-based image segmentation. To answer this question, we recorded visually evoked potentials, both in man and in awake monkey, to a stimulus specifically designed to signal the presence of neuronal activity related to contour from motion processing. We report here that response components specific to contour from motion were elicited only when the stimulus yielded a contour percept. In awake monkey, the sources of these components were located within the supra- and infra-granular layers of primary visual cortex. We conclude that V1 is involved in image segmentation processing.

We designed a stimulus consisting of moving random dot patterns, arranged in a checkerboard fashion (Fig. 1). The dots were moved in two sequences of four apparent motion phases which were repeated constantly. Motion in the two sets of dots was in opposite directions in the odd-numbered stimulus phases and in the same directions in the even-numbered stimulus phases, whereas all directions of motion were distributed evenly over both odd and even phases. Any difference between the average responses to odd and even phases thus cannot come from local motion signals but indicates a contribution from mechanisms that are sensitive to relative motion. By triggering the averaging procedure to each odd-numbered stimulus phase, this difference is expressed in the odd harmonic content of the evoked responses. The 'checksizes' of the stimuli were varied from 4 × 4 pixels per check up to a single check filling the whole screen (256 × 256 pixels). The latter stimulus provides a control for the equality of the two stimulus phases with respect to local modulations and should not evoke odd harmonics.

Figure 2 shows scalp-recorded evoked potentials to the full range of stimuli, for both man and awake monkey. Odd harmonics are present in the responses to 'checksizes' of 128, 64, 32 and 16 pixels, indicating that relative motion-sensitive mechanisms contribute to these responses. We also asked our human subjects to view single sequences of the stimuli and to report whether they perceived a checkerboard or not. The results

* To whom correspondence should be addressed.

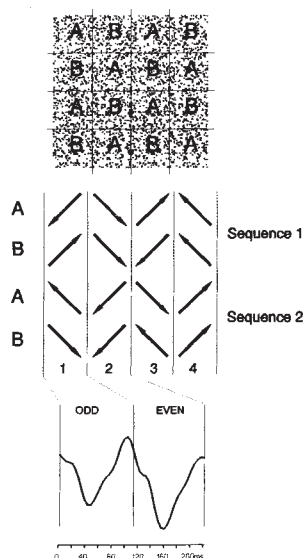


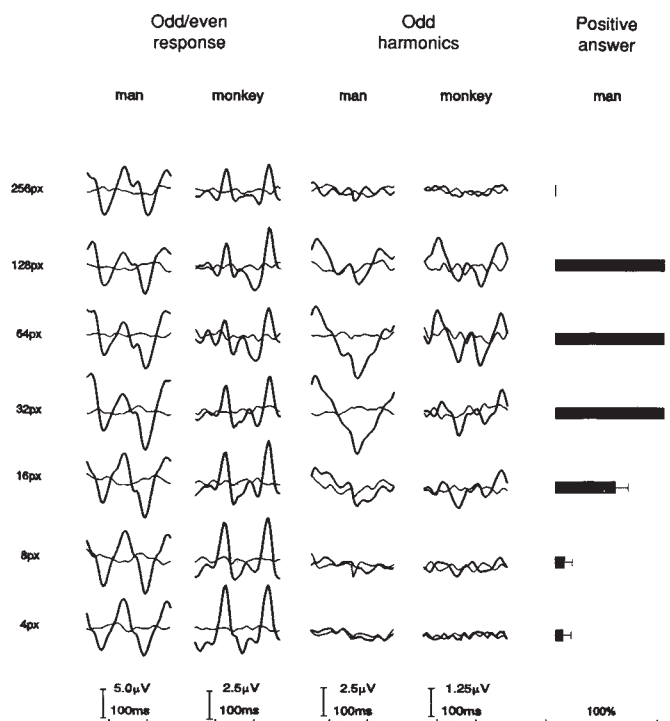
FIG. 1 The stimulus models. Top: a CRT monitor screen, with a 256×256 pixel resolution and subtending about 6.7° , is divided into two sets of squares A and B, in a checkerboard fashion. Each square is filled with randomly chosen black and white pixels, with equal probability. For the sake of illustration, dot density is smaller in the figure, and lines have been added to demarcate the squares. Middle: every 110 ms the pixels within each square are coherently displaced over a distance of 1 pixel in one out of four oblique directions. Two stimulus sequences are used. Each sequence consists of 4 different phases, numbered 1–4. The arrows indicate the direction of apparent motion for each of the 4 phases, and for the two sets of pixel-filled checks (A or B). When, for example, movement in one set of checks (A) is, as in phase 1 of sequence 1, left-downward, and in the other set (B) right-upward, a clear contour from motion checkerboard is perceived (odd numbered phases), which is not visible with both sets of pixels moving in the same direction (even phases). Note that the moving pixels never cross their checkerboard boundary. This means that in the odd as well as the even phases, pixels appear and disappear along the checkerboard boundaries. Bottom: evoked potentials are averaged by triggering on each odd stimulus phase. Numerous repetitions of sequence 1 are followed by an equal amount of sequence 2 repetitions, followed again by sequence 1, and so on. Thus the odd phase of the response is the average of the responses during phases 1 and 3 of the two sequences; the even phase of the responses is the average of the responses during phases 2 and 4 of the two sequences. Responses to the two sequences, which proved to be identical, are averaged. All directions of motion, and other low-level cues, contribute equally to the odd and even response halves. Because motion contrast is only present at the odd stimulus phases, any difference between the odd and even response halves, that is the presence of any odd harmonics, signals a contribution to the response from mechanisms sensitive to relative motion.

are shown in the rightmost column of Fig. 2. From the comparison of columns 3 and 5 of Fig. 2, it is evident that odd response harmonics are only present when subjects have the percept of a motion-induced checkerboard. This indicates that the odd harmonics reflect activity which is related to contour from motion processing. Because the stimulus design assures that odd harmonics can only be generated by a mechanism sensitive to relative motion, the results of Fig. 2 show furthermore that these odd harmonics are generated by a mechanism that uses relative motion to compute contours and not by a mechanism which responds to the local discontinuities of motion directions. Such a mechanism would have produced the largest

responses for the smaller checksizes, where such discontinuities are most numerous.

To localize the neuronal origin of the odd response harmonics, we recorded from awake monkeys, with a linear array of 16 electrodes, perpendicular to the cortical surface of the operculum of area 17. Current source density analysis of the measured spatio-temporal potential profile yields a reliable estimate of the distribution of current sources and sinks underlying the responses^{6,7}. These currents are mainly generated by synaptic activity⁷. Figure 3 shows the results for a stimulus with a 'check-size' of 64 pixels. The leftmost column shows the 16 responses recorded from the positions indicated by the cortical lamination

FIG. 2 Evoked potentials psychophysical responses to the stimuli. Responses for 7 different 'checksize's. Checksize refers to the size of the pixel-filled checks for a set of stimuli where screen size, pixel density and modulation are all as described in Fig. 1. In man, we recorded from a mid-occipital lead (Oz) just above the inion and in monkey from a position overlying the operculum of area 17, both against a mid-frontal reference. Monkeys were trained to fixate the stimulus monitor¹⁷. The heavy traces depict the averaged responses, and the thin traces the noise estimates obtained by $\pm$ averaging¹⁸. Positivity is upwards. The left two columns show the actual responses, columns 3 and 4 their odd harmonic contents, which reveal a contribution from motion contrast sensitive mechanisms. Odd harmonics were calculated by subtracting the first half of the response from the second half, and the second half from the first half. The rightmost column depicts data from a psychophysical experiment on man. It shows percentages of positive answers, for 7 subjects, when forced to report whether they saw a checkerboard or not, upon the display of single sequences of the stimuli. Error bars indicate standard deviations ($N=7 \times 10$ for each checksize used).



given at the centre. The second column gives the distribution of current sources (upward deflections) and sinks (downwards), underlying these potentials. A complicated pattern emerges, with activity in all layers. The third column shows the odd harmonics of the responses of column one. The pattern of sources and sinks related to this potential distribution (column 4) reveals contour from motion processing activity in the supra-granular layers and in layer 5. Layer 4C activity is absent, which strengthens the notion that the odd harmonics signal a specific, intracortical process.

Our results show that signals related to contour from motion processing are already present in primary visual cortex. In a set of experiments described elsewhere, very similar results were obtained for texture segregation based on orientation differences. Again, segmentation-specific components were only evoked when orientation differences elicited the percept of texture segregation^{8,9} and intracortical response sources were found in layers 2/3 and 5 of primary visual cortex⁹. We conclude therefore that primary visual cortex is involved in the process of image segmentation in general. Moreover, our results particularly suggest the specific involvement of the layer 2/3 and layer 5 neurons, which are reciprocally connected with extensive lateral spread¹⁰. Of course we cannot exclude on the basis of the present data the possibility that other brain areas are involved as well. Higher

visual areas might also contribute to the segmentation process in V1 by means of feedback projections. The classic notion that V1 merely provides a local analysis of the visual field, and that global processes are the privilege of higher areas seems, however, no longer tenable.

Single unit studies have never demonstrated neurons in V1 to respond in a feature detection kind of way to relative motion or texture segregation borders^{11,12}. It has been shown, however, that responses to stimuli inside the receptive field are often modulated by simultaneous stimulation of the receptive field surround¹³⁻¹⁵. Our results suggest that this context modulation is a reflection of activity which represents image segmentation at the level of the neuronal ensemble. This favours population coding models of image segmentation¹⁶, instead of hierarchical models. □

Received 7 December 1992; accepted 30 March 1993.

1. Julesz, B. *Foundations of Cyclopean Perception* (University of Chicago Press, Chicago, 1971).
2. Braddick, O. J. *Vision Res.* **14**, 519 (1974).
3. Van Essen, D. C. in *Cerebral Cortex 3, Visual Cortex* (eds Peters, A. C. & Jones, E. G.) (Plenum, New York, London, 1985).
4. Nakayama, K., *Vision Res.* **25**, 427-433 (1985).
5. Maunsell, J. H. R. & Newsome, W. T. *A. Rev. Neurosci.* **10**, 363-401 (1987).
6. Nicholson, C. & Freeman, J. A. *J. Neurophysiol.* **38**, 356-368 (1975).
7. Mitzdorf, U. *Physiol. Rev.* **65**, 37-100 (1985).
8. Lamme, V. A. F., Van Dijk, B. W. & Spekreijse, H. *Vision Res.* **32**, 797-807 (1992).
9. Lamme, V. A. F., Van Dijk, B. W. & Spekreijse, H. *Visual Neurosci.* (in the press).
10. Blasdel, G. G., Lund, J. S. & Fitzpatrick, D. *J. Neurosci.* **5**, 3350-3369 (1985).
11. Hildreth, E. C. & Koch, C. *A. Rev. Neurosci.* **10**, 477-533 (1987).
12. Nothdurft, H. C. & Li, C. Y. *Vision Res.* **25**, 99-113 (1985).
13. Allman, J., Miezin, F. & McGuinness, E. *A. Rev. Neurosci.* **8**, 407-430 (1985).
14. Orban, G. A., Gulyas, B. & Vogels, R. *J. Neurophysiol.* **57**, 1792-1812 (1987).
15. Gilbert, C. D. & Wiesel, T. N. *Vision Res.* **30**, 1689-1701 (1990).
16. Sporns, O., Tononi, G. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 129-133 (1991).
17. Dagnelie, G., Spekreijse, H. & Van Dijk, B. W. *Vis. Neurosci.* **3**, 509-525 (1989).
18. Schimmel, H. *Science* **157**, 92-94 (1967).
19. Kraut, M. A., Arezzo, J. C. & Vaughan, H. G. Jr *Electroenceph. Clin. Neurophysiol.* **62**, 300-312 (1985).

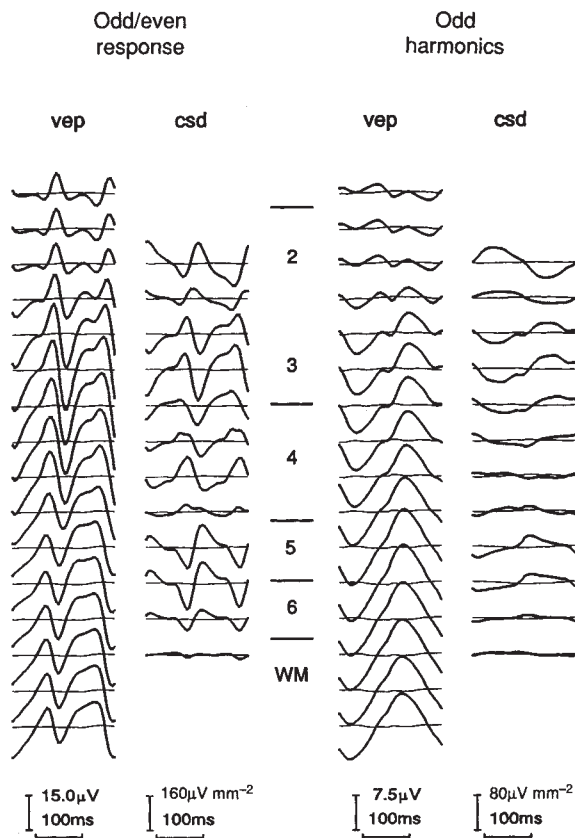


FIG. 3 Primary visual cortex response sources. Intracortical responses from awake macaque parafoveal V1 to a stimulus with a 'checksize' of 64 pixels. Three monkeys were trained to fixate the stimulus monitor¹⁷ and similar results were obtained from all of them. Interelectrode distance is 150 μ m. Column 1 gives the actual responses and column 2 the current source density (CSD) profile⁶ underlying these responses. Column 3 gives the odd harmonic contents of the responses of column one and column 4 their CSD profile. Heavy lines give the averaged responses and thin lines the noise estimates. Cortical lamination is indicated between columns 2 and 3. The laminar positions of recordings were obtained by electrical lesioning, and subsequent histological examination and by comparing flash VEPs to homogeneous field flashes with a repetition rate of 2 Hz recorded at the beginning of each session, with previously published data¹⁹.

Peripheral nerve injury triggers noradrenergic sprouting within dorsal root ganglia

Elsbeth M. McLachlan*, Wilfrid Jänig†, Marshall Devor‡ & Martin Michaelis†

* Department of Physiology & Pharmacology, University of Queensland, Queensland 4072, Australia

† Physiologisches Institut, Christians-Albrechts Universität, Kiel 2300, Germany

‡ Department of Cell and Animal Biology, Life Sciences Institute, Hebrew University, Jerusalem 91904, Israel

IN humans, trauma to a peripheral nerve may be followed by chronic pain syndromes which are only relieved by blockade of the effects of sympathetic impulse traffic¹⁻⁴. It is presumed that, after the lesion, noradrenaline released by activity of sympathetic postganglionic axons excites primary afferent neurons by activating α -adrenoceptors^{2,5}, generating signals that enter the 'pain pathways' of the central nervous system. The site of coupling is unclear. In some patients local anaesthesia of the relevant peripheral nerve⁶ does not alleviate pain, implying that ectopic impulses arise either within the central nervous system, or in proximal parts of the primary afferent neurons. In experimentally lesioned rats, activity can originate within the dorsal root ganglia^{7,8}. Here we report that, after sciatic nerve ligation, noradrenergic perivascular axons in rats sprout into dorsal root ganglia and form basket-like structures around large-diameter axotomized sensory neurons; sympathetic stimulation can activate such neurons repetitively. These unusual connections provide a possible origin for abnormal discharge following peripheral nerve

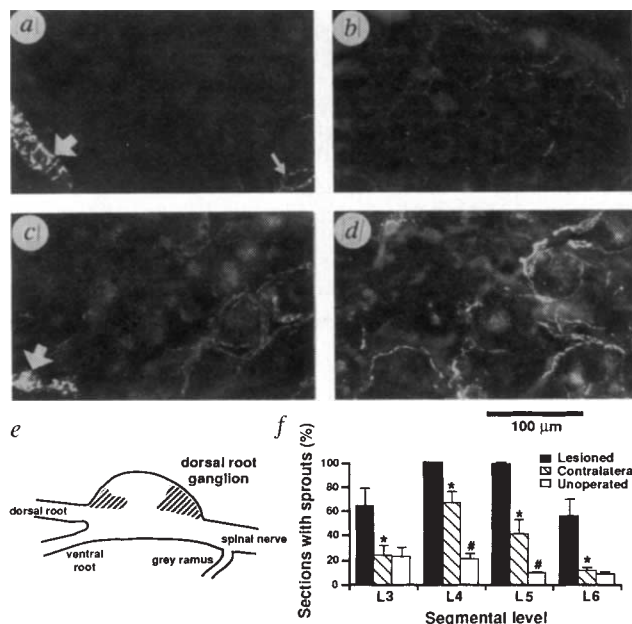


FIG. 1 *a-d*, Photomicrographs showing catecholamine-fluorescent axons in (a) right and (b) left L5 DRGs from an animal 30 days after section of left sciatic nerve, and left DRGs of other animals (c) 50 days and (d) 70 days post-operation (p.o.). Large arrows (a, c), perivascular plexuses; small arrow (a), axons near capsule. Scale bar in *d* applies to all micrographs. *e*, Main locations of sprouts (hatched) in lesioned DRGs. *f*, Frequency of sprouting in DRGs ≥ 30 days p.o. (defined as the percentage of sections containing any axons (other than perivascular) penetrating deeper than one cell-diameter into the ganglion. Note that this greatly underestimates the extent of sprouting in lesioned DRGs). Bars indicate s.e.m., $n=5-7$ ganglia. Significant differences ($P < 0.05$, t -test) between lesioned (black) and contralateral (hatched) ganglia (*); contralateral and unoperated (white) ganglia (#). There were no sex or lateral differences in unoperated animals.

METHODS. Wistar rats (74 ± 5 days old, 180–330 g, 6 male, 8 female) were anaesthetized (ketamine 60 mg kg^{-1} ; xylazine 10 mg kg^{-1} , injected i.p.) and, aseptically, left sciatic nerves were ligated and cut at mid-thigh level with excision of about 5 mm of the distal stumps. Two males and two females (58 ± 3 days old) served as unoperated controls. After 2–70 days, animals were deeply anaesthetized (pentobarbitone, 80 mg kg^{-1} , i.p.), and perfused with oxygenated physiological saline. DRGs L3–6 and paravertebral ganglia L3–5 were removed, together with spinal nerves and sciatic nerves at the level of the lesion. Paired lesioned and contralateral tissues were rapidly frozen in moulds, and serial ($20 \mu\text{m}$) cryostat sections processed using glyoxylic acid^{16,27}. Direct demonstration of catecholamine was preferred to immunohistochemistry for transmitter-synthesizing enzymes or neuropeptides because non-catecholaminergic neurons may express these substances after lesions^{28,29}. All axonal structures are 'catecholamine-fluorescent' unless specified otherwise. Some animals autotomized their two lateral toes, but there was no correlation between this behaviour and histology.

damage. Further, in contrast to the sprouting of intact nerve terminals into nearby denervated effector tissues in skin^{9,10}, muscle¹¹, sympathetic ganglia¹² and sweat glands¹³, the axons sprout into a target which has not been partially denervated.

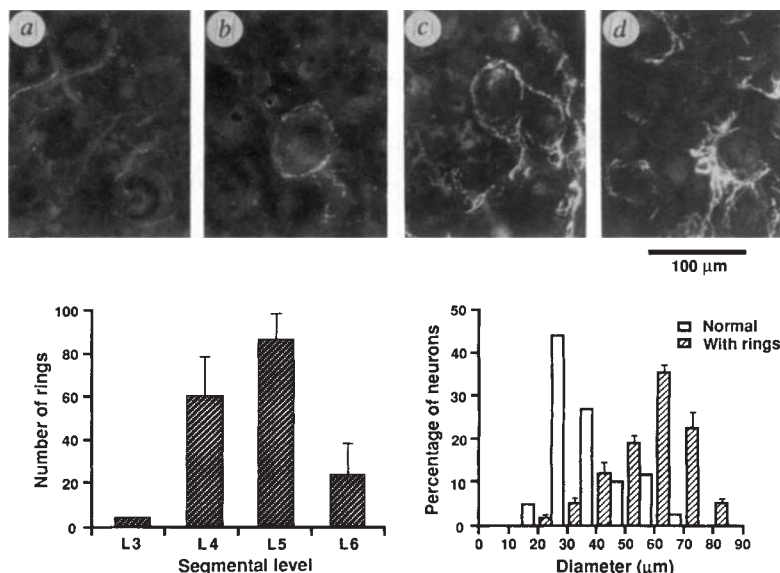
We have investigated the distribution in rats of noradrenergic terminals in dorsal root ganglia (DRGs) and sympathetic ganglia projecting in sciatic nerves which had been ligated and transected up to 10 weeks previously. Within the DRGs from unoperated animals and on sides contralateral to the lesions, catecholamine-containing axons form varicose plexuses around small-diameter ($15-50 \mu\text{m}$) blood vessels (Fig. 1*a*). They are occasionally found near the perimeter of the DRGs¹⁴; single fine non-varicose axons rarely penetrate deeper than two soma diameters.

Our most startling finding was that many fine fluorescent axons penetrated those DRGs (L3–6) that contained somata of lesioned sciatic axons (Figs 1*b-d* and 2*a-d*). Brightness and

extent of these sprouts increased with time after the nerve lesion. At 10–18 days, fine largely non-varicose axons could not be photographed before fading. After 3 weeks, thin axons spread among the ganglion cells and bright varicose terminals were wrapped around some somata (Figs 1*c, d*, 2*d*, and 3*a-d*), preferentially located near the poles (Fig. 1*e*). The presence of many fine axons after 10 weeks suggested that sprouting was ongoing (Fig. 1*d*). Axons penetrated contralateral DRGs more frequently than unoperated controls but far less than lesioned DRGs (Fig. 1*f*).

In the fluorescence microscope, some DRG somata on the lesioned side had a dark outer zone and an eccentric nucleus, which is typical after axotomy, with more orange lipofuchsin granules than in normal neurons (Figs 1, 2). Axons encircled some of these presumed axotomized neurons forming 'rings' (Figs 1*c, d* and 2*a-d*); a small proportion (<5%) of these rings consisted of clusters of varicosities encapsulating a cell (Figs

FIG. 2 *a-d*, Photomicrographs of neurons surrounded by 'rings' (defined as fluorescent axons running closely adjacent to a somatic profile for $\geq 35\%$ of its perimeter) in lesioned DRGs from animals (a) 30, (b) 50 and (c, d) 70 days p.o. Calibration (d) applies to all micrographs. *e*, Numbers of 'rings' in DRGs on the lesioned side of 4 animals 50 days p.o. Bars indicate s.e.m. (except for L3, $n=2$). No attempt has been made to correct for double counting; overestimation on the basis of soma size could be as much as 60%, but this would be countered by the limited proportion of soma surface associated with most rings. *f*, Diameters of L5 DRG neuronal profiles surrounded by rings (hatched columns; $n=265$, 26–102 profiles from each of 4 animals; bars indicate s.e.m.), together with normal soma diameter distribution for L5 (white columns, data for resin-impregnated sections from ref. 30). Diameters were determined directly with the aid of an eyepiece micrometer; as well as including offcuts of large-diameter neurons, values may be overestimates because of the absence of neuronal staining. No differences in the ringed populations were detected between L4 and L5 ganglia, or between sexes.



1c, 2b-d). Cells bearing rings rarely abutted and had no preferential location. Seven weeks after the lesion, 132-199 rings (mean $\pm$ s.e.m., 160 ± 14 ; $n = 4$) were distributed between L3-6 ganglia, depending on the rostrocaudal fixation of the lumbosacral plexus (Fig. 2e); the number had doubled by 10 weeks (326 , $n = 1$). Rings were rare in DRGs without axotomized somata; there were 2 in 16 DRGs from unoperated animals, and 0-11 (mean, 2.3 ± 1.0) in 18 DRGs contralateral to lesioned sciatic nerves. Rings surrounded some small neuronal profiles deep in the ganglion but clearly preferred the more superficial large-diameter somata (Fig. 2f).

The sympathetic neurons that projected in the lesioned sciatic nerve are unlikely to be those that sprouted into the DRGs. The fine axons visible in normal peripheral nerve trunks¹⁵ disappeared within a few days after the lesion and, other than rare thick axons near the DRGs, which resembled regenerating axons¹⁶, virtually no noradrenergic axons were detected in the lesioned nerves except in scar tissue adjacent to the neuroma¹⁷. Many paravertebral neurons lost their fluorescence on the lesioned side. These data suggest that most ligated sympathetic axons degenerate retrogradely, lose their capacity to synthesize noradrenaline and/or die¹⁸. Instead, it was clear from tracing individual sprouts through serial sections that most sprouts were derived from intact perivascular plexuses (Fig. 3; ref. 19). When the left L4 grey ramus (projecting to spinal nerve L5) was sectioned at its origin at the time of one sciatic ligation, all noradrenergic structures were absent in the ipsilateral L5 DRG.

Sprouting of perivascular axons not directly affected by the distal lesion, rather than injured postganglionic neurons, is consistent with the sprouting observed in contralateral DRGs. The latter also indicates that sprouting of intact sympathetic neurons was probably not induced by a transneuronal signal within paravertebral ganglia. Changes in contralateral sensory²⁰ and motor²¹ axons after nerve lesions have been explained by crossed intraspinal signals. Here, 'rings' were preferentially formed around large-diameter axotomized DRG somata, suggest-

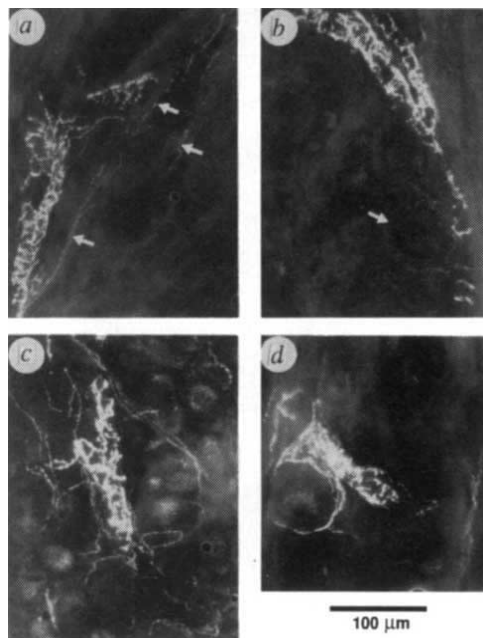


FIG. 3 a-d, Photomicrographs of perivascular plexuses and sprouts in DRGs from the lesioned side. a, Fine sprouts (arrows) in the spinal roots running towards the DRG (50 days p.o.). b, Very fine sprouts (arrow) surrounding the soma of an axotomized neuron (arrow, 30 days p.o.). c, Sprouts spreading out among DRG cells (70 days p.o.). d, Perivascular axons forming a ring around an adjacent large diameter soma (70 days p.o.). Calibration (d) applies to all micrographs.

ing these neurons might trigger sprouting by release of some chemotactic substance.

Functional coupling between the newly formed 'synapses' and the DRG cells has been demonstrated by activation of some myelinated primary afferent axons at sites proximal to sciatic neuromas in rats with similar lesions after repetitive stimulation of the sympathetic outflow. The discharge began after the end of a stimulus train and lasted for many seconds (Fig. 4b); it was not initiated within the neuroma (Fig. 4c). Such responses were abolished by phentolamine (intravenous injection of $10-30 \mu\text{g kg}^{-1}$; Fig. 4d). Noradrenaline elicits slow depolarizations of some peripheral²² and central^{23,24} neurons via α_1 -adrenoceptors, which is linked to a reduction in resting potassium conductance. This would increase neuronal excitability and potentially elicit discharge.

These aberrant noradrenergic terminals around DRG somata could be responsible for sympathetically mediated paraesthesias as a result of initiating afferent activity during ongoing and reflex sympathetic postganglionic discharge. Pain might be generated if central nociceptive pathways become sensitized by activity in low-threshold mechanosensitive (large-diameter) afferents, as in secondary hyperalgesia³. Alternatively, central terminals of large-diameter DRG cells sprout to fill vacated synaptic sites in lamina II of the dorsal horn after sciatic lesions²⁵. Thus, activity in large DRG cells initiated by release

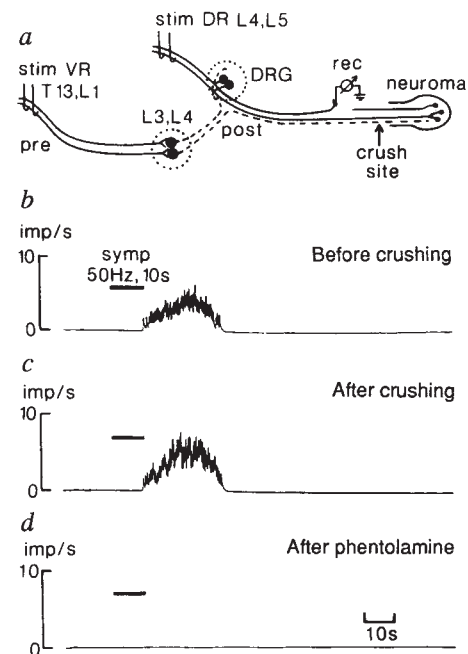


FIG. 4 a, Diagram of experimental arrangement (Sabra strain of Wistar rat, 470 g, male; anaesthetized with pentobarbital 50 mg kg^{-1} i.p., supplemented by 20 mg kg^{-1} as needed). The rat was paralysed with gallamine triethiodide and artificially ventilated; 11 days after sciatic nerve ligation and transection (see legend to Fig. 1), all left dorsal and ventral roots T13-L5 were cut. Extracellular recording from a single myelinated afferent fibre in lesioned sciatic nerve (identified by stimulation of dorsal roots (DR) L4 and L5, conduction velocity 24.8 m s^{-1}). b, c, This afferent axon showed no ongoing activity. Stimulation of preganglionic axons (pre) in ventral roots (VR) T13 and L1 (0.5 ms, 5 mA, 10-s trains at 50 Hz at bars) was followed by a burst of discharge lasting about 30 s. This stimulus strength was maximal for activation of a postganglionic (post) volley recorded from a branch of an intact muscle nerve in the upper thigh. The response was not affected by crushing the nerve between the neuroma and the recording site with jeweller's forceps for 15 s, controlled visually under the stereomicroscope. d, The response was abolished after administration of phentolamine ($20 \mu\text{g kg}^{-1}$, injected i.v.). Activity was integrated with a ratemeter (time constant, 0.5 s).

of noradrenaline on to their somata might be converted into nociceptive signals in the central nervous system. Nerve lesions can provoke bilateral hyperalgesia which is alleviated by sympathetic blockade²⁶, consistent with contralateral sprouting as observed here. However, unlike the central sprouting of distally lesioned sensory neurons, the noradrenergic terminals of undamaged neurons sprout into a target without synaptic sites. □

Received 2 November 1992; accepted 5 April 1993.

1. Bonica, J. J. in *The Management of Pain* (ed. Bonica, J. J.) 220–243 (Lea and Febinger, Philadelphia, 1990).
2. Jänig, W. in *Pain and the Sympathetic Nervous System* (ed. Stanton-Hicks, M.) 17–89 (Kluwer, Boston, 1990).
3. Jänig, W. & Schmidt, R. F. (eds) *Reflex Sympathetic Dystrophy. Pathophysiological Mechanisms and Clinical Implications* (VCH, Weinheim and New York, 1992).
4. Willis, W. D. (ed.) *Hyperalgesia and Allodynia* (Raven, New York, 1992).
5. Jänig, W. & Koltzenburg, M. in *Towards a New Pharmacotherapy of Pain* (eds Basbaum, A. I. & Besson, J. M.) 331–352 (1991).
6. Sunderland, S. *Nerves and Nerve Injuries* (Churchill Livingstone, Edinburgh, 1978).
7. Wall, P. D. & Devor, M. *Pain* **17**, 321–339 (1983).
8. Devor, M. & Wall, P. D. *J. Neurophysiol.* **64**, 1733–1746 (1990).
9. Devor, M., Schonfeld, D., Seltzer, Z. & Wall, P. D. *J. comp. Neurol.* **185**, 211–220 (1979).

10. Diamond, J., Holmes, M. & Coughlin, M. *J. Neurosci.* **12**, 1454–1466 (1992).
11. Brown, M. C., Holland, R. L. & Hopkins, W. G. *A. Rev. Neurosci.* **4**, 17–42 (1981).
12. Maehlen, J. & Nja, A. *J. Physiol., Lond.* **319**, 555–567 (1981).
13. Kennedy, W. R. & Sakuta, M. *Ann. Neurol.* **15**, 73–78 (1984).
14. Kummer, W., Gibbins, I. L., Stefan, P. & Kapoor, V. *Cell Tissue Res.* **261**, 595–606 (1990).
15. Small, J. R., Scadding, J. W. & Landon, D. N. *J. neural. Sci.* **100**, 98–107 (1990).
16. Jobling, P., McLachlan, E. M., Jänig, W. & Anderson, C. R. *J. Physiol., Lond.* **454**, 107–128 (1992).
17. Goldstein, R. S., Raber, P., Govrin-Lippmann, R. & Devor, M. *Neurosci. Lett.* **94**, 58–63 (1988).
18. Jänig, W. & McLachlan, E. M. *J. comp. Neurol.* **225**, 302–311 (1984).
19. Crutcher, K. A. *Brain Res. Bull.* **9**, 501–508 (1982).
20. Kolston, J., Lisney, S. J. W., Mulholland, M. N. C. & Passant, C. D. *Neurosci. Lett.* **130**, 187–189 (1991).
21. Rotschker, S. *Trends Neurosci.* **11**, 363–366 (1988).
22. Akasu, T., Gallagher, J. P., Nakamura, T., Shinnick-Gallagher, P. & Yoshimura, M. *J. Physiol., Lond.* **361**, 165–184 (1985).
23. Fukuda, A., Minami, T., Nabekura, J. & Oomura, Y. *J. Physiol., Lond.* **393**, 213–231 (1987).
24. Yoshimura, M., Polosa, C. & Nishi, S. *Brain Res.* **414**, 138–142 (1987).
25. Woolf, C. J., Shortland, P. & Coggeshall, R. E. *Nature* **355**, 75–78 (1992).
26. Shir, Y. & Seltzer, Z. *Pain* **45**, 309–320 (1991).
27. de la Torre, J. C. *J. Neurosci. Meth.* **3**, 1–5 (1980).
28. Wakisaka, S., Kajander, K. C. & Bennett, G. J. *Neurosci. Lett.* **124**, 200–203 (1991).
29. Morris, J. L. & Gibbins, I. L. *Int. J. dev. Neurosci.* **7**, 521–531 (1989).
30. Schmalbruch, H. *Anat. Rec.* **219**, 323–329 (1987).

ACKNOWLEDGEMENTS. We thank M. Bauer for animal care and for histochemistry, and J. Keast, P. Jobling and P. Sah for comments on the manuscript. This work was funded by the National Health & Medical Research Council of Australia. Physiological experiments (W.J., M.D. and M.M.) were supported by the German-Israeli Foundation for Scientific Research and Development.

The Nagel anomaloscope and seasonal variation of colour vision

G. Jordan & J. D. Mollon

Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

IN 1948 the German physicist, Manfred Richter, reported that colour vision has a seasonal variation^{1,2}. For four colour-normal subjects, he found a sinusoidal variation in the proportion of red and green required to match a monochromatic yellow, the equation known as the 'Rayleigh match'³. In summer, subjects required more red in their mixture. The measurements were made with the Nagel anomaloscope, an instrument introduced in 1907^{4,5} and which today, essentially unchanged, remains the definitive clinical instrument for classifying the many phenotypic variations in colour vision. The variation that Richter recorded in the red–green ratio was large (three Nagel units), and it now takes on fresh interest because it is comparable in size to the difference in Nagel settings later reported between normal observers of different genetic types^{6,7}. We have been able to replicate Richter's result, but report here that it is almost certainly instrumental: the Nagel anomaloscope proves to be very sensitive to ambient temperature.

One of us (G.J.) made almost daily Rayleigh matches for over one year using a conventional modern Nagel anomaloscope run from a stabilized power supply. Five settings were made at about the same time each day. G.J.'s matches (Fig. 1c) show seasonal variation with similar phase to Richter's data: a greater proportion of long-wave light is required in the match during the summer months. The amplitude of the variation in our data, if expressed in terms of anomalous quotients⁸ (to allow comparison of the different instruments), is close to that of Richter's.

In the December of this study, we noticed that Nagel matches dropped suddenly at weekends and at the beginning of the Christmas holiday, when the heating of our laboratory was turned off. We suspected that Richter's effect might be mediated by variation in ambient temperature. We do not know the indoor temperatures in his laboratory in 1946–1947, but it is likely, particularly in summer, that they reflected in part the variation in outdoor temperature, and in Fig. 1b we plot the monthly outdoor temperatures for Berlin for that period (supplied by the Berlin meteorological office). In Fig. 1d we show analogous data recorded at the National Institute of Agricultural Botany, Cambridge, during the period of our own study. The four functions of Fig. 1 are similar: sine waves fitted to the data agree

in phase to within a few degrees. But these are only correlational data, and season has many correlates; so we manipulated experimentally room temperature and then the local temperature of the Nagel anomaloscope.

The Nagel anomaloscope is essentially a spectroscope, consisting of an ocular, a compound direct-vision prism, and three entrance slits that define the primaries (Fig. 2). Its modern form, the Model I manufactured by Schmidt & Haensch, is optically similar to the two versions of the instrument described by Nagel.

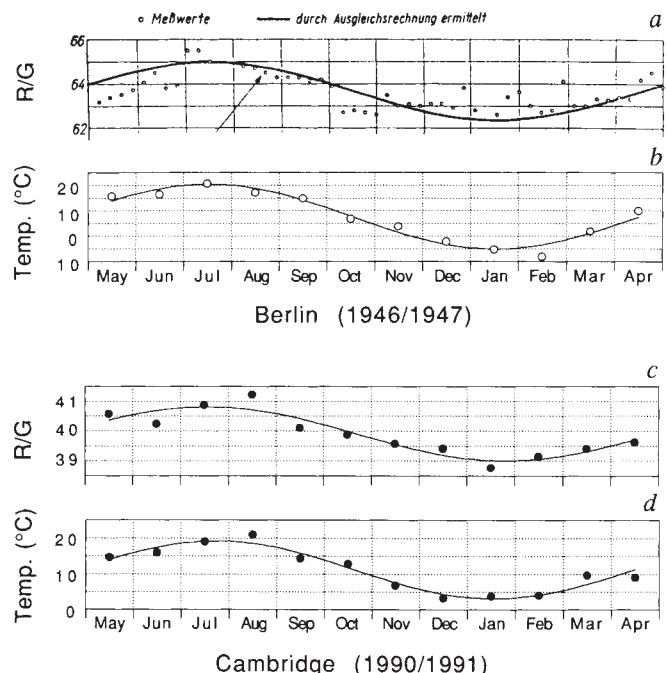


FIG. 1 Seasonal variations in Rayleigh matches. a, Rayleigh matches of four normal male observers obtained in Berlin, May 1946 to April 1947¹; c, mean Rayleigh matches of one female observer recorded in Cambridge, May 1990, to April 1991¹⁵. The Rayleigh match represents the relative proportions of red and green light required in a mixture to match monochromatic yellow light, and is expressed in the scale units of the anomaloscopes used in the two experiments. b, d, Outdoor temperatures recorded in Berlin and Cambridge during the corresponding periods. The solid curves shown in panels b–d are best-fitting sine waves, with period and phase free to vary.

We have examined three instruments: two in Cambridge, a Model I supplied in 1988 and a Model II from early this century; and a 1974 Model I in Dortmund.

Figure 3a shows, for two colour-normal observers, how settings on a Model I vary with room temperature. There is a correlation of >0.9 between Rayleigh match and ambient temperature: the higher the temperature, the more protan the setting, that is, the greater the proportion of red light required in the mixture. The slope amounts to 0.175 Nagel units per degree Celsius. A very similar dependency is seen for the older, Model II, instrument, and for the Dortmund instrument (Fig. 3b, c).

How does ambient temperature affect Nagel settings? There are six possibilities. (1) The change might be in the observer. For example, there might be a change in body temperature, which in turn altered the slopes of the long-wave limbs of the photopigments⁹. Or there might be a change in the ionic environment of the photoreceptors that changed their spectral sensitivities¹⁰. (2) A variation in the power supply could lead to a variation in lamp current and thus in the colour temperature of the output. (3) Ambient temperature might directly influence the lamp's operating temperature. (4) Temperature might affect the position or width of the entrance slits of the anomaloscope, and thus the effective primaries. The entrance slits that define the red and green primaries are 8.5 mm apart and the relative proportions of the two primaries are controlled by a movable tongue (Fig. 2b). The maximum width of either slit is less than 0.5 mm. Because a distance of only 8.5 mm corresponds to a wavelength difference of over 100 nm and because the Rayleigh match is made at a red-green balance where the normal eye is very sensitive, a very small change in the position or relative widths of the slits could be expected to be visible¹¹. (5) Thermo-mechanical stress may change the angle that the ocular or the objective makes to the prism and thus change the effective primaries. (6) Temperature may change the refractive index of the prism and thus the effective wavelengths of the primaries. The temperature coefficient for the refractive index of typical optical glass is $5 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$. Because the refractive index typically varies with wavelength by $9 \times 10^{-5} \text{ nm}^{-1}$, the shift in wavelength would be $0.05 \text{ nm } ^\circ\text{C}^{-1}$ (ref. 12). As there is a large variation in temperature coefficients and the prism in the Nagel is likely to be a compound of flint and crown glass, it is difficult to estimate the exact effect, but the change is large enough to be visually apparent.

The following experiments were designed to distinguish

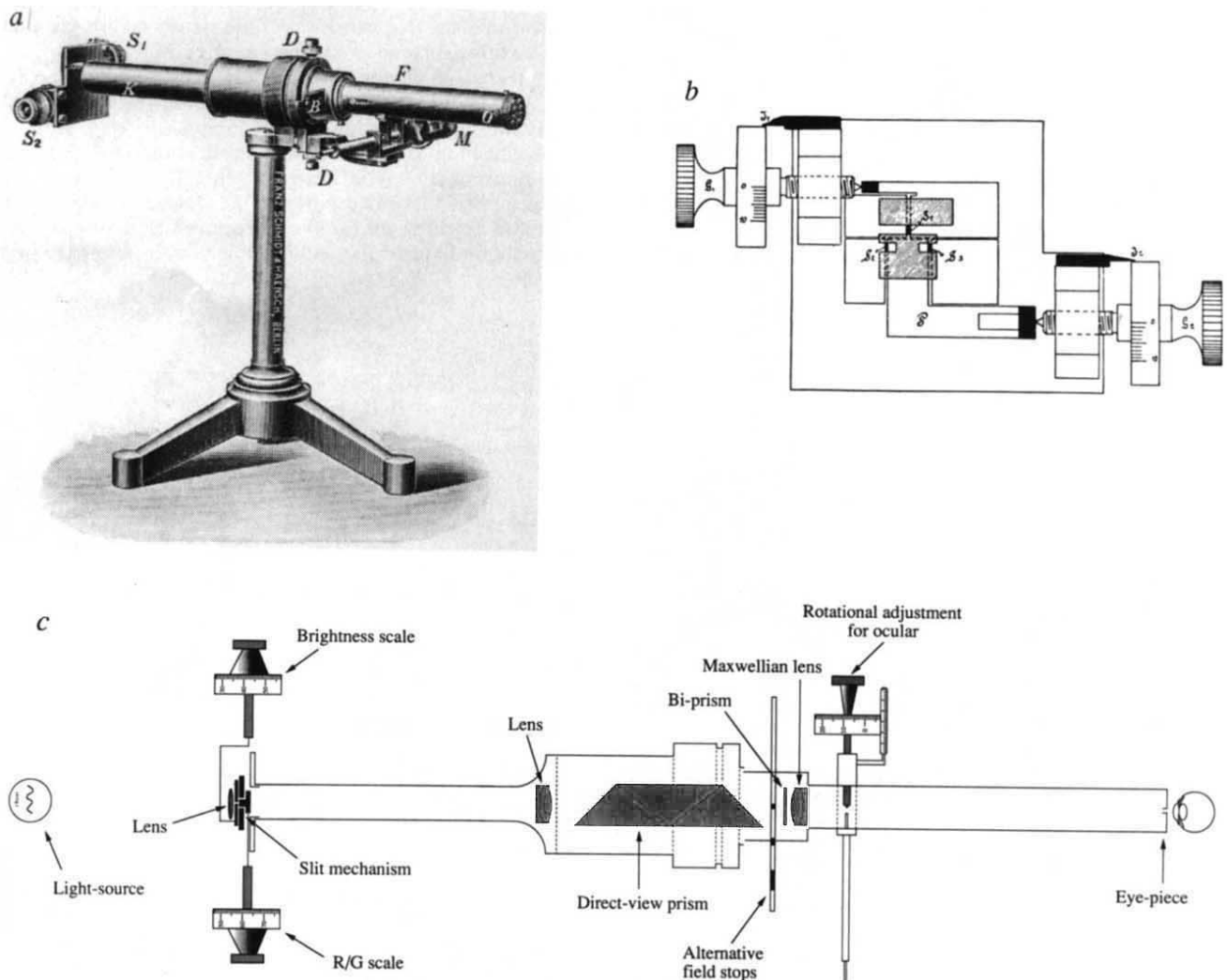


FIG. 2 a, Nagel anomaloscope Model II in its original form⁴. The eyepiece is on the right of the instrument and the slit mechanism on the left. The difference between Models I and II is that Model II has a facility to change the angle of the ocular, and thus the three primaries selected for the equation. b, Detail of the slit mechanism. Movement of the tongue (P) between the two sides of the slit increases either the red or the green

primary while reducing the other. The upper slit controls the radiance of the yellow primary. c, Plan view of the Nagel anomaloscope, Model II. The direct-vision prism is housed in the centre part of the instrument, and a biprism and lens provide a horizontally divided field seen in Maxwellian view. The optics of the modern Model I anomaloscope are very similar¹⁶.

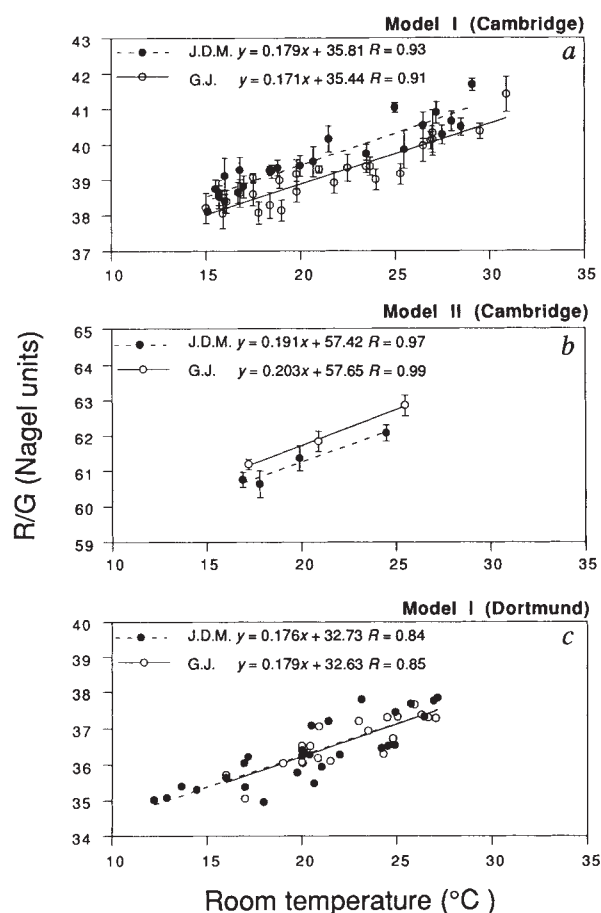


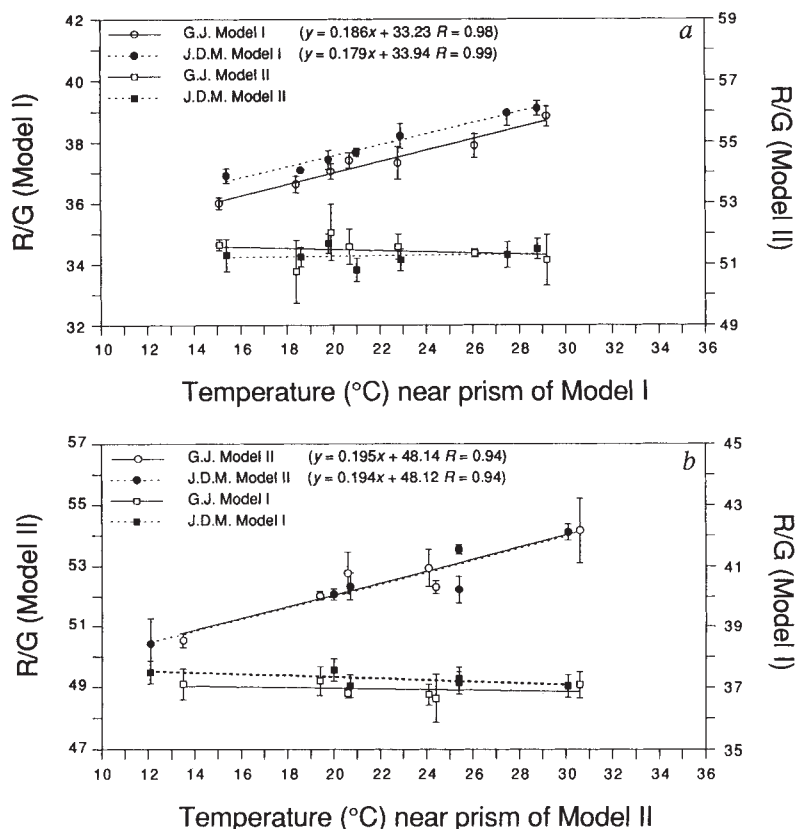
FIG. 3 Variation of Rayleigh matches as a function of room temperature. R/G Nagel settings are shown for two normal observers on three different instruments. The two Cambridge instruments were run from a stabilized power supply. *a, c*, Measurements for the two different Model I instruments: the slope of the function is ~ 0.17 Nagel units per degree C. *b*, Settings on an antique Model II instrument manufactured by Spindler and Hoyer: the slope of the function is about 0.19 Nagel units per degree C. For all three instruments the rate of change in anomalous quotient is about 0.01 per degree C.

between these possibilities. We set up our Model I and our Model II anomaloscopes side by side and operated them from the same power supply, voltage-stabilized to within a variation of less than 1%. The laboratory was air-conditioned, and room temperature was constant within a range of 1 °C. The two observers made measurements alternately on the two instruments while we locally heated or cooled the prism housing of one of the two. Figure 4*a* shows the Rayleigh settings for the case where the Model I was varied and the Model II served as control; Fig. 4*b* shows the complementary experiment. In both cases, a variation in Rayleigh match is seen only for the manipulated instrument.

These experiments allow us to rule out three of the possibilities listed. The variation cannot be in the observers, who remained in a constant environment and made settings alternately on the two instruments. Nor can it lie in the power supply, common to the two instruments. And the variation is unlikely to be in the lamps, which were turned on only long enough to do each set of five matches, and we measured little thermal coupling between the prism housing and the lamp housing.

Direct heating of the slits had less effect than heating of the prism housing, but it is not possible fully to uncouple the temperature of these two parts of the instrument (and of the ocular and objective tubes). So we removed from our Model II instrument both the ocular and the objective, isolating the prism

FIG. 4 Experiment in which the prism housing of one anomaloscope was gently heated or cooled, while the second instrument was held at a constant temperature. Settings were made alternately on the test and on the control instrument. In each panel the left-hand ordinate, and the upper two sets of data points, represent the R/G settings of the manipulated anomaloscope, whereas the right-hand ordinate and the lower two sets of data points represent the corresponding settings on the control instrument. In each case, the abscissa represents the temperature measured (with a K-type thermocouple) at the surface of the prism housing of the manipulated instrument. The data points for the manipulated instrument show slopes of ~ 0.17 – 0.19 Nagel units per degree C, but no systematic changes are seen for the control instrument.



in its mount. We passed a He-Ne laser beam through the prism from the ocular face to the objective face, and monitored its displacement on a distant wall as room temperature was manipulated. Part of the beam was reflected at the first (ocular) face of the prism, allowing us to monitor mechanical movement of the prism and mount. The laser source was in a stable environment. An increase of 15 °C in the ambient temperature at the prism had no effect on the reflected beam, but the refracted beam was displaced by 0.53 mrad away from the red entrance slit, implying that the primaries vary with temperature.

We are satisfied that Rayleigh matches on the Nagel anomaloscope exhibit a thermal dependence. The effect is instrumental and probably arises from change in refractive index at the air-prism interfaces or at interfaces within the compound prism. In the absence of any evidence that Richter controlled the ambient temperature of his instrument, we must conclude that his seasonal variation is largely or wholly artefactual. Doubt must similarly fall on the correlation reported by Waaler for Nagel settings of parents and children⁶, because family members are likely to have been tested at similar times; and although later molecular studies have confirmed a genetic source of variation in normal Rayleigh matches¹³, temperature must have been a significant source of variance in all population studies that have used the Nagel anomaloscope.

It may seem unlikely that the Nagel anomaloscope has been used for so long without anyone noticing its thermal dependence, but we end by citing a passing caution made in 1915: "A further

influence on the investigation, and one that is sometimes detectable and uncontrollable, may not be generally known, namely the temperature of the apparatus [Nagel anomaloscope] or of the experimental room. With it the dispersion-power of the prisms of the apparatus changes noticeably, and therewith the wavelengths of the monochromatic light that is used"¹⁴. □

Received 4 January; accepted 31 March 1993.

1. Richter, M. Z. *wiss. Photogr. Photophys. Photochem.* **43**, 209–237 (1948).
2. Richter, M. *Klin. Mon. Augenheilkunde* **119**, 561–575 (1951).
3. Rayleigh, L. *Nature* **25**, 64–66 (1881).
4. Nagel, W. A. Z. *Augenheilkunde* **18**, 201–222 (1907).
5. Nagel, W. in *Handbuch der Physiologischen Methodik* (ed. Tigerstedt, R.) (Hirzel, Leipzig, 1909).
6. Waaler, G. H. M. *Nature* **215**, 406 (1967).
7. Waaler, G. H. M. *Avhandl. Det Norske Videnskaps-Akad. Oslo I Mat.-Naturv. Klasse. Ny Serie, No.* 9, 1–25 (1967).
8. Pokorny, J., Smith, V. C., Verriest, G. & Pinckers, A. J. L. G. *Congenital and Acquired Color Vision Defects* (Grune & Stratton, New York, 1979).
9. De Vries, H. *Experientia* **4**, 357 (1948).
10. Knowles, A. *Vision Res.* **20**, 475–485 (1980).
11. Moreland, J. D. in *Colour Vision Deficiencies II* (ed. Verriest, G.) 14–18 (Karger, Basel, 1974).
12. Merton, T. R. *Proc. R. Soc. B* **113**, 704–708 (1927).
13. Winderickx, J. et al. *Nature Genet.* **1**, 251–255 (1992).
14. Köllner, H. *Arch. Augenheilkunde* **78**, 302–335 (1915).
15. Jordan, G. *Polymorphism of Normal Colour Vision in Humans* (University of Cambridge, Cambridge, 1992).
16. Heinsius, E. *Farbsinnstörungen und ihre Prüfung in der Praxis* (Ferdinand Enke, Stuttgart, 1973).

ACKNOWLEDGEMENTS. We thank C. R. Cavonius for advice and for laboratory facilities in Dortmund, The National Institute of Agricultural Botany and the Wetteramt, Berlin for temperature records, and J. Bennett, J. Moreland, M. Webster and P. Whittle for discussion. Supported by the MRC and the British Council.

Intercellular signalling in *Drosophila* segment formation reconstructed *in vitro*

Susan Cumberledge & Mark A. Krasnow

Department of Biochemistry, School of Medicine, Stanford University, Stanford, California 94305–5307, USA

GENETIC studies show that intercellular signalling is involved in key steps in *Drosophila melanogaster* development, but it has not previously been possible to investigate these processes in simplified *in vitro* systems. Analysis of *engrailed* (*en*) and *wingless* (*wg*) and other segment polarity genes suggests that two or more intercellular signalling processes may be involved in intrasegmental patterning^{1–3}. Expression of *en* and *wg* begins about three hours after egg laying, in adjacent rows of cells in the posterior half of each segmental primordium^{4–7}. In *wg*[–] embryos and in conditional mutants in which *wg* function is inactivated during a critical period between three and five hours after egg laying, early *en* expression begins normally but then disappears within several hours^{4,8–10}. The *wg* gene encodes a protein highly similar to the product of the mouse *Wnt-1* proto-oncogene, a secreted glycoprotein^{11,12}; *wg* protein is proposed to function as an extracellular signal, maintaining *en* expression and activating other molecular and morphogenetic processes in nearby cells^{4,8,9,13}. Several lines of evidence support the model, including the secretion of *wg* protein in the embryo^{7,14}, genetic mosaic experiments^{15–17} and cell lineage studies¹⁸. We tested this model using purified embryonic cells isolated by whole animal cell sorting¹⁹, and validated three key predictions: (1) when *en*-expressing cells from early embryos are grown alone in culture, they rapidly and selectively lose *en* expression; (2) purified *wg*-expressing cells provide a locally active signal that prevents this loss; (3) heterologous cells engineered to express *wg* also show signalling activity, indicating that *wg* protein alone, or in conjunction with more generally expressed factors, is the signal.

TABLE 1 S2 cells expressing *wg* protein rescue *en* expression

Culture conditions	Fraction of EN-cells expressing <i>en</i> protein after 3 h (%)	Enhancement
EN-cells + S2hsWG(–) cells	9.1	
EN-cells + S2hsWG(+) cells	33.2	3.6 ×

The *wg*-expressing cell line S2hsWG(+) was constructed by co-transfecting S2 cells with pV91.1 DNA, in which the *D. melanogaster hsp70* heat-shock promoter is fused to a 2.9-kb *wg* cDNA minigene (M. van den Heuvel and R. Nusse, personal communication), and pMK33 DNA, a plasmid carrying the *hygro* gene of *Escherichia coli*, which confers resistance to hygromycin. The control cell line S2hsWG(–) was constructed in parallel using pV91.2 DNA, which is identical to pV91.1 except that the *wg* cDNA is inserted in the antisense orientation (M. van den Heuvel and R. Nusse, personal communication). Stable transformants were selected by continuous growth in SM+hygromycin at 0.2 mg ml^{–1}. Untransformed S2 cells and S2hsWG(–) cells did not express *wg* protein before or after a 45-min heat shock at 37 °C, whereas S2hsWG(+) cells synthesized significant levels of *wg* protein after heat shock, as detected by western blotting (data not shown). Induced S2hsWG(+) cells did not express detectable levels of *en* protein (data not shown). To test for rescue activity, S2hsWG(+) cells and S2hsWG(–) cells were plated at ~70% confluence in separate culture wells and incubated at 25 °C. After 12–16 h, the cells were heated to 37 °C for 45 min and then returned to 25 °C for 60 min to allow *wg* protein to accumulate. Purified EN-cells were then added to each culture and incubation continued. After 3 h, the number of EN-cells still expressing *en* protein in the culture was determined, expressed as a percentage of EN-cells added to the culture. The number of EN-cells added to each culture was determined by sorting an equal number of EN-cells directly onto a polylysine coated slide, and then fixing immediately and staining for *en* protein. For six independent trials, the average effect was 3.6 (±1.4)-fold.

To test the model, *en*-expressing cells (EN-cells) were isolated by fluorescence-activated cell sorting (FACS) of cells from 3–5-hour embryos carrying an *en* promoter-*lacZ* (β -galactosidase) transgene, after staining with a fluorogenic β -galactosidase substrate. The whole animal cell sorting (WACS) purification procedure has been described¹⁹. (Fig. 1 legend). If early EN-cells

are cultured alone, *en* expression should quickly switch off as the cells would be deprived of the signal from their *wg*-expressing neighbours. To test this, the purified EN-cells were grown in Schneider's *Drosophila* medium for various times, and then fixed and stained with an anti-*en* antibody. Immediately after isolation nearly all the purified cells expressed *en* protein, but the fraction of cells expressing *en* protein rapidly declined during culturing, and by four hours only ~10 per cent of the cells continued to express detectable levels (Fig. 1).

The decline in *en* protein levels was not due to cell death or generalized protein degradation. Cell viability determination by trypan-blue exclusion demonstrated that >99% of the EN-cells remained viable throughout the culture period, and >95% continued to express high levels of two control proteins (Fig. 1). Thus in the absence of other cell types there is a rapid and selective decrease in the synthesis or stability of *en* protein.

We next tested whether purified *wg*-expressing cells (WG-cells) could prevent loss of *en* expression. WG-cells were isolated by WACS from 3–5-hour-old embryos carrying a *wg* promoter-

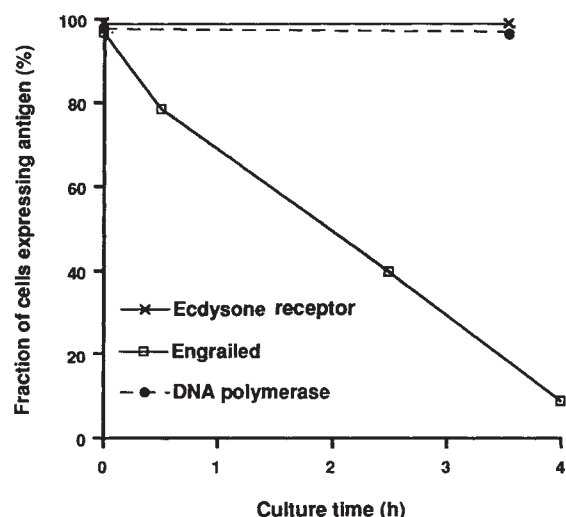


FIG. 1 Selective loss of *en* expression in purified embryonic *en*-expressing cells (EN-cells). EN-cells were purified by WACS from 3–5-hour-old embryos and grown in culture for the times indicated. Cells were then fixed and expression of *en* protein, the EcR ecdysone receptor protein, and the large subunit of DNA polymerase- α was determined by immunocytochemistry. Data for the different proteins are from separate experiments.

METHODS. EN-cells were isolated as described¹⁹, except that cells were washed only once in Schneider's medium (SM) before filtration through Nitex. Briefly, dissociated embryonic cells, prepared from 3–5-h-old 2.5EPLHT + 5.0R embryos, were stained with a fluorogenic β -galactosidase substrate and two fluorescent cell viability dyes. Live cells with high β -galactosidase activity were purified by FACS. This strain harbours a third chromosome insert of a P element containing a 7.5-kb *en* promoter fragment driving *lacZ* expression, similar to the 7.5-kb constructs of Hama *et al.*²¹ (C.-N. Chen and T. Kornberg, personal communication). At this time during embryogenesis, expression of β -galactosidase parallels endogenous *en* expression (C.-N. Chen and T. Kornberg, personal communication; and our unpublished results), and the purified cells also express high levels of *en* protein¹⁹. For each time point, equal numbers of live EN-cells ($5\text{--}20 \times 10^3$ cells depending on the experiment) were sorted into 0.1 ml SM in a polyllysine-coated glass chamber slide (6.5-mm wells; Nunc) and cultured at 25 °C. (The cell numbers given are the nominal number sorted by the FACS instrument, but for reasons that are unclear the number of sorted cells determined by microscopy was consistently significantly smaller.) After culturing, cells were fixed for 12 min in 4% paraformaldehyde, collected, immunostained and viewed as described²¹. The primary antibodies were: *en* monoclonal antibody (mAb) 4D9 (1:2,500 dilution of ascites fluid), EcR mAb DDA2 and DNA polymerase mAb 512 (1:30 and 1:3,000 dilutions of hybridoma culture supernatants, respectively). Random fields were scored for expression of the antigen. Similar results were obtained when EN-cells were cultured on slides precoated with extracellular matrix from *D. melanogaster* S2 cells.

lacZ transgene, and then co-cultured with EN-cells. Under these conditions, a substantial fraction of the cells still expressed *en* antigen after three to four hours in co-culture (Fig. 2a, and data not shown). Although the final level of expressing cells varied in different experiments, cultures containing WG-cells consistently gave 3–4.5-fold more *en*-expressing cells than parallel cultures of EN-cells grown alone. The stimulation was specific to WG-cells. There was no rescue of *en* expression when EN-cells were cultured either with purified embryonic cells that do not express the *wg-lacZ* transgene, or with S2 cells²⁰, a cell line derived from *Drosophila* embryos (Table 1). There was also specificity in the responding cells, as *en* expression was not aberrantly activated in purified WG-cells or in S2 cells exposed to the signal (Table 1).

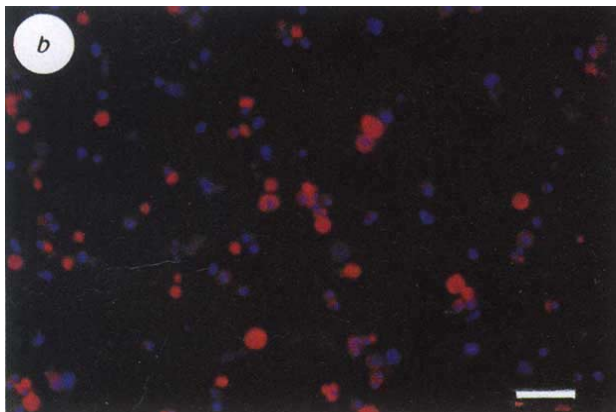
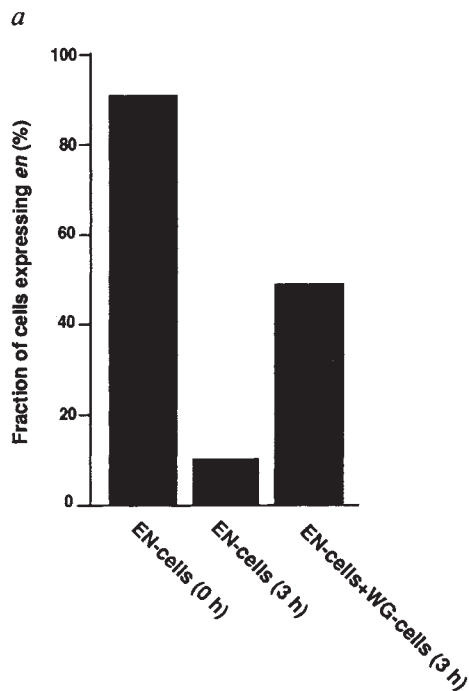
Using time-lapse microscopy, we observed that ~10% of the cells underwent one or more rounds of division during the culture period, and few if any cells lysed, regardless of whether EN-cells and WG-cells were grown together or separately (data not shown). Together with the trypan blue data on cell viability already described, these results indicate that WG-cells do not affect EN-cell survival or proliferation in the assay. Rather, WG-cells must stimulate the expression or stability of *en* protein in EN-cells.

If *wg* protein is the signal, and it is the only signal provided by the WG-cells, then heterologous cells expressing *wg* protein should also have signalling activity. A stably transfected S2 cell line, S2hsWG(+), which expresses *wg* protein under heat-shock promoter control, and a non-expressing control cell line, S2hsWG(–), were grown overnight and then heat-shocked to induce expression of the transgenes. EN-cells were added to

TABLE 2 Signalling activity is blocked by a membranous filter and is not present in conditioned medium from WG-cells

Culture conditions	Fraction of cells expressing <i>en</i> protein after 3–4 h (%)	Enhancement
a, Cells separated by filter		
EN-cells	11.6	
EN-cells/WG-cells	10.0	0.9 ×
b, Cells on same side of filter		
EN-cells	16.8	
EN-cells + WG-cells	54.0	3.2 ×
c, Cells separated by filter		
EN-cells/S2hsWG(–)	23.4	
EN-cells/S2hsWG(+)	25.5	1.1 ×
d, Conditioned medium (CM)		
EN-cells + S2hsWG(–) CM	7.9	
EN-cells + S2hsWG(+)	10.7	1.3 ×

a, Methods as for Fig. 2, except cells were grown in a Costar transwell dish with a 6.5-mm diameter upper chamber and a 8.5-mm lower chamber which are separated by a 0.4 μ m microporous polycarbonate membrane. In one culture the upper chamber contained EN-cells (10^5) in 0.2 ml Schneider's medium (SM) and the bottom chamber contained 0.2 ml SM; in another the upper chamber contained EN-cells (5×10^4) in 0.2 ml SM and the lower chamber contained WG-cells (5×10^4) in 0.2 ml SM. For two independent trials, the average enhancement was 1.1 (± 0.3)-fold. b, Conditions as for a, except that when both EN-cells and WG-cells were present they were grown together in the upper chamber and the bottom chamber contained 0.2 ml SM. c, Conditions as for a, except that S2hsWG(+) or S2hsWG(–) cells were grown in the lower chamber instead of embryonic WG-cells. Growth of S2 cell lines and heat induction are described in Table 1 legend. For three independent trials, the average effect was 1.3 (± 0.3)-fold. d, Methods as for Table 1, except that 0.1 ml conditioned medium from the indicated cell lines was used instead of cells. For seven independent trials, the average effect was 1.4 (± 0.3)-fold. Conditioned medium from S2hsWG(+) cells and S2hsWG(–) cells was prepared by plating the cells at ~70% confluence and growing them in SM at 25 °C for 12–16 h and then at 37 °C for 45 min, followed by recovery at 25 °C for 60 min to induce expression of the introduced heat-shock gene. Cells were then removed by centrifugation and the supernatant added immediately to freshly purified EN-cells.



the cultures, and incubations continued for three hours before assay of *en* expression. Many more EN-cells continued to express *en* in cultures containing S2hsWG(+) cells than in parallel cultures with S2hsWG(-) cells (Table 1). The magnitude of rescue was similar to that with embryonic WG-cells. Thus, *wg* protein is the only cell-type-specific factor required for signalling, and *wg* protein alone, or in conjunction with more generally expressed factors provided by S2 cells, constitutes the signal.

Under standard co-culture conditions, a significant fraction of the EN-cells are close to or contacting WG-cells (Fig. 2b). To test whether the signal acts locally or is freely diffusible, co-culture experiments were done with the signalling and responding cells in two chambers separated by a 0.4- μ m filter. This should prevent contact between the two cell types but allow passage of many proteins and other small signalling molecules. Under these conditions, there was little or no rescue of *en* expression (Table 2a and c). Control experiments in which the two cell types were cultured on the same side of the filter gave the usual level of rescue (Table 2b). Conditioned medium from *wg*-expressing S2hsWG(+) cells was also tested in the standard assay and found to have little activity (Table 2d). Together, these results indicate that the signal is locally active, either because it is highly unstable or not freely diffusible. This is probably true *in vivo* as well, as *wg* activity and protein appear

FIG. 2 Loss of *en* expression can be rescued by co-culture with embryonic WG-cells. **a**, Purified EN-cells were cultured alone or with an equal number of purified WG-cells. Cultures were grown for the indicated times and *en* expression measured. For samples containing both EN-cells and WG-cells, the fraction of total cells expressing *en* protein was determined and then doubled to obtain the fraction of EN-cells expressing *en*. For five independent trials, the average effect was 3.7 (± 0.6)-fold. In no experiment did we observe rescue of *en* expression in all EN-cells; this could be because the signal is limiting under the culture conditions, or because some of the purified EN-cells are incompetent to respond or require another signal. It may be possible to achieve higher levels or complete rescue with more highly synchronized embryonic cells. **b**, Distribution of EN-cells and WG-cells under co-culture conditions. EN-cells were stained with a cytoplasmic dye (CMTMR: (5-(and-6-)-((4-chloromethyl)benzoyl)amino)tetramethylrhodamine), and both cell types were counterstained with DAPI (4',6-diamidino-2-phenylindole). EN-cells appear red or pink in the photomicrograph and WG-cell nuclei are blue. Scale bar, 25 μ m. Note that the figure indicates the initial distribution of cells in the culture; for technical reasons, we have not been able reliably to determine the final distribution of cells expressing *en* protein after culturing because the cells are not highly adherent and many do not remain attached to the substratum during antibody staining.

METHODS. WG-cells and EN-cells were purified from 3–5-h-old embryos by WACS. WG-cells were isolated from W2, a strain harbouring a third chromosome insert of a P element containing a 2-kb *wg* promoter fragment driving *lacZ* expression (P. Lawrence, personal communication); the transgene is expressed like endogenous *wg* at this time in development. Cells (10^5 EN-cells, or 5×10^4 each of WG-cells and EN-cells) were cultured as for Fig. 1, except that slides were pretreated with S2 cells for deposition of extracellular matrix. For the pretreatment, S2 cells were plated at ~70% confluence, grown for 16–20 h at 25 °C and removed by washing with SM. In preliminary experiments, we observed higher levels of rescue when co-culturing was in pretreated wells rather than on polylysine-coated slides. For **b**, cells were treated as described, except that the crude EN-cell preparation was incubated in SM + 80 nM CMTMR (Molecular Probes) for 5 min at 25 °C before filtration and washing. After staining by the standard procedure (except propidium iodide was omitted in the EN-cell purification because of spectral overlap with CMTMR), WG-cells and EN-cells (5×10^4 each) were sorted directly into 0.1 ml SM on a polylysine-coated slide. DAPI was added to 0.2 mg ml⁻¹. Cells were centrifuged briefly to attach them to the slide and then fixed in 4% paraformaldehyde for 5 min. The sample was washed with phosphate-buffered saline, mounted in 20% glycerol in PBS, and photographed.

to be restricted to a region extending only one or at most a few cell diameters away from cells that express messenger *wg* RNA^{7,14,18}.

Our *in vivo* experiments demonstrate directly that intercellular signalling occurs between WG-cells and EN-cells during segment formation and establish several essential features of the process. First, *en* expression in early EN-cells is in an unstable state, and their intrinsic program is to turn off *en* expression selectively shortly after it initiates. Second, neighbouring WG-cells provide a locally active signal that prevents this turn-off. No other cell types are required for this effect, nor is the integrity of the parasegment or parasegment borders. Third, the signal does not appear to function as a mitogen or trophic factor for EN-cells. And fourth, *wg* protein alone may be the signal. If other segment polarity gene products are required in WG-cells for signalling, then they must also be present in S2 cells, or their requirement is somehow by-passed *in vitro*. The *in vitro* system should allow biochemical analysis of the signal and the signal transduction process, and help elucidate the roles of other segment polarity genes in the signalling pathway^{1–3}. □

Received 16 November 1992; accepted 16 March 1993.

1. Ingham, P. W. *Curr. Opin. Genet. Dev.* **1**, 261–267 (1991).
2. DiNardo, S. & Heemskerk, J. *Sem. Cell Biol.* **1**, 173–183 (1990).
3. Ingham, P. W. & Martinez Arias, A. *Cell* **68**, 221–235 (1992).

4. Martinez Arias, A., Baker, N. E. & Ingham, P. W. *Development* **103**, 157–170 (1988).
5. Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. *Cell* **40**, 45–53 (1985).
6. Baker, N. E. *EMBO J.* **6**, 1765–1773 (1987).
7. van den Heuvel, M., Nusse, R., Johnstone, P. & Lawrence, P. A. *Cell* **59**, 739–749 (1989).
8. DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassiss, J. A. & O'Farrell, P. H. *Nature* **332**, 604–609 (1988).
9. Bejsovec, A. & Martinez Arias, A. *Development* **113**, 471–485 (1991).
10. Heemskerk, J., DiNardo, S., Kostriken, R. & O'Farrell, P. *Nature* **352**, 404–410 (1991).
11. Rijsewijk, F. et al. *Cell* **50**, 649–657 (1987).
12. Nusse, R. & Varmus, H. E. *Cell* **69**, 1073–1087 (1992).
13. Riggelman, R., Schedl, P. & Wieschaus, E. *Cell* **63**, 549–560 (1990).
14. Gonzalez, F., Swales, L., Bejsovec, A., Skaer, H. & Martinez Arias, A. *Mech. Dev.* **35**, 43–54 (1991).
15. Morata, G. & Lawrence, P. A. *Dev. Biol.* **56**, 227–240 (1977).
16. Baker, N. E. *Dev. Biol.* **125**, 96–108 (1988).
17. Wieschaus, E. & Riggelman, R. *Cell* **49**, 177–184 (1987).
18. Vincent, J. P. & O'Farrell, P. *Cell* **68**, 923–931 (1992).
19. Krasnow, M. A., Cumberledge, S., Manning, G., Herzenberg, L. A. & Nolan, G. P. *Science* **251**, 81–85 (1991).
20. Schneider, I. J. *Embryol. exp. Morph.* **27**, 353–365 (1972).
21. Hama, C., Ali, Z. & Kornberg, T. B. *Genes Dev.* **4**, 1079–1093 (1990).

ACKNOWLEDGEMENTS. We thank M. van den Heuvel and R. Nusse for discussion and for *wg* expression vectors and antibodies; P. A. Lawrence and the late C. Cabrera for the *wg-lacZ* strain; C.-N. Chen and T. B. Kornberg for the *en-lacZ* strain; W. Talbot and D. Hogness for *EcoR* antibodies; the I. R. Lehman laboratory for DNA polymerase antibodies; and R. Nusse, J. Klingensmith, L. Fradkin and the members of our laboratory for comments on the manuscript. S.C. was supported by an NIH postdoctoral fellowship and M.A.K. is a Lucille P. Markey Scholar in Biomedical Science. This work was supported by grants from the Lucille P. Markey Charitable Trust and NIH to M.A.K.

A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation

Michael T. Michalek^{*†}, Ethan P. Grant^{*†},
Colette Gramm^{*†}, Alfred L. Goldberg[‡]
& Kenneth L. Rock^{*†}

^{*} Division of Lymphocyte Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA

[†] Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡] Department of Cellular and Molecular Physiology, Harvard Medical School, Boston, Massachusetts 02115, USA

THE degradation of most cellular proteins starts with their covalent conjugation with ubiquitin^{1,2}. This labels the proteins for rapid hydrolysis to oligopeptides by a (26S) proteolytic complex containing a (20S) degradative particle called the proteasome^{3,4}. Some system in the cytosol also generates antigenic peptides from endogenously synthesized cellular and viral proteins^{5–10}. These peptides bind to newly synthesized class I major histocompatibility complex molecules in the endoplasmic reticulum and peptide/class I complexes are then transported to the cell surface for presentation to cytotoxic T cells^{11,12}. How these peptides are produced is unknown, although a modification that promotes ubiquitin-dependent degradation of a viral protein enhances its presentation with class I¹³ and indirect evidence suggests a role for proteolytic particles closely resembling and perhaps identical to the proteasome^{4,12,14,15}. Using cells that exhibit a temperature-sensitive defect in ubiquitin conjugation, we report here that non-permissive temperature inhibited class I-restricted presentation of ovalbumin introduced into the cytosol, but did not affect presentation of an ovalbumin peptide synthesized from a minigene. These results implicate the ubiquitin-dependent proteolytic pathway in the production of antigenic peptides.

Ubiquitin-protein conjugation and protein degradation are reduced in ts20 cells at nonpermissive temperature (41 °C)^{16,17} because of irreversible inactivation of a mutant E1 enzyme. E1 catalyses the ATP-dependent activation of ubiquitin, an essential first step in conjugation of ubiquitin to cellular proteins^{1,2}. The ts20 and E36 (wild-type) cells were transfected with H-2K^b and ICAM-1 complementary DNAs to make them appropriate antigen-presenting cells (APCs) for the murine T-cell hybridoma RF33.70 (ref. 18). Presentation of ovalbumin (OVA) by ts20.10.2 (mutant) (Fig. 1a) but not E36.12.4 (wild-type) (Fig. 1b) cells

was greatly decreased by preincubating the cells at 41 °C. These APCs acquired ovalbumin by pinocytosis⁸. The uptake of extracellular fluid (0.94 ± 0.27 nl per 10⁶ cells at 41 °C versus 0.96 ± 0.47 nl per 10⁶ cells at 37 °C, mean ± s.d. of four experiments) by mutant cells was not inhibited as determined by the uptake of radiolabelled poly(vinylpyrrolidone) (PVP), a commonly used marker of fluid-phase pinocytosis¹⁹. APCs were incubated with PVP and ovalbumin under identical conditions. Thus, the failure of mutant cells to present ovalbumin was not due to a defect in pinocytosis.

Incubation at 41 °C did not affect the ability of mutant (Fig. 1c) or wild-type (Fig. 1d) cells to present exogenous OVA_{257–264} peptide (which represents the naturally processed epitope from ovalbumin, residues 257–264). When added to the extracellular medium, the synthetic OVA_{257–264} peptide binds directly to H-2K^b on the cell surface^{20,21}. Therefore, presentation of OVA peptide-K^b complexes at the APC surface was not significantly affected by blocking ubiquitin conjugation.

We next examined the effects of 41 °C on the synthesis of H-2K^b. After 1 h incubation at 41 °C or 37 °C, APCs were radiolabelled with [³⁵S]methionine, and H-2K^b molecules were sequentially immunoprecipitated with a monoclonal antibody (Y-3)²² specific for assembled K^b molecules followed by a rabbit anti-heavy chain antiserum (exon 8)²³. Similar amounts of assembled H-2K^b were immunoprecipitated from mutant cells incubated at the two temperatures (Fig. 2a). Similar amounts

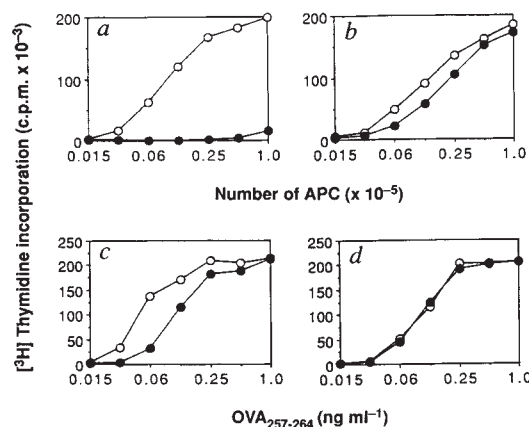


FIG. 1 MHC class I-restricted presentation of ovalbumin by ts20.10.2 and E36.12.4. Ts20.10.2 (a, c) and E36.12.4 (b, d) cells were incubated at 41 °C (filled circles) or 37 °C (open circles) and then incubated with (a, b) or without (c, d) OVA. The indicated number of APCs incubated with OVA were cultured with RF33.70 (a, b), or control APCs (5 × 10⁴) were cultured with RF33.70 and the indicated concentration of OVA_{257–264} peptide (c, d). **METHODS.** Ts20.10.2 and E36.12.4 (2 × 10⁶ cells per ml) were incubated for 1 h at 41 °C or 37 °C and then OVA (20 mg ml⁻¹) or buffer (control APCs) was introduced into the cytosol (10 min at 37 °C) by osmotic lysis of pinosomes as previously described^{8,30}. The washed APCs (2 × 10⁶ cells per ml) were subsequently incubated for 1 h at 37 °C to allow for the expression of processed antigen on the cell surface and then fixed with paraformaldehyde (1%) as previously described³⁰. Incubations during and after the introduction of OVA were done at 37 °C to eliminate temperature as a variable between the groups while the APCs internalized and processed the antigen. During this 37 °C incubation, some recovery of ubiquitin conjugating activity (16 ± 10%, mean ± s.d. of three experiments) did occur in the 41 °C-treated mutant cells. The fixed APCs were cocultured with the RF33.70 T cell hybrid which produces interleukin-2 (IL-2) on recognition of OVA_{257–264} + K^b on the surface of APCs¹⁷ and IL-2 was quantified as previously described³⁰. E36 and ts20 cells were cotransfected with H-2K^b (provided by G. Wanek and subcloned into pcDL-SRα296neo³¹) and ICAM-1 (in pH₂Apr-1-neo, provided by S. Hedrick and A. Brian³²) cDNAs using the Lipofectin reagent (BRL) as previously described³³. G418-resistant clones were screened for expression of ICAM-1 and H-2K^b with antibody YN1/1.7.4 (ref. 33) and B8-24.3 (ref. 34), respectively. The transfected cell lines ts20.10.2 and E36.12.4 were selected and passaged at 31 °C and 37 °C, respectively.

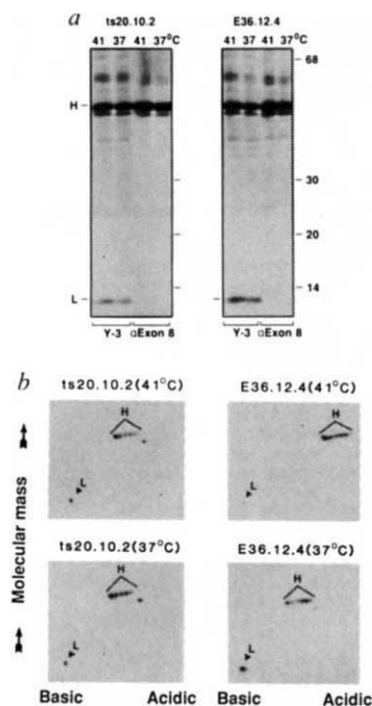


FIG. 2 Effect of temperature on the synthesis and maturation of H-2K^b. *a*, Ts20.10.2 and E36.12.4 cells were incubated at 41 °C or 37 °C and then metabolically radiolabelled. H-2K^b was sequentially immunoprecipitated with Y-3 antibody (ref. 35) followed by anti-exon 8 (ref. 23) antiserum (from B. Barber, University of Toronto) and analysed by one-dimensional SDS-PAGE. Y-3 binds only assembled H-2K^b molecules, whereas the anti-exon 8 antiserum binds both assembled H-2K^b molecules and free heavy chains. The position of H-2K^b heavy chains (H) and β_2 -microglobulin (L) are indicated on the left. Relative molecular mass standards ($M_r \times 10^{-3}$) are shown to the right. *b*, Ts20.10.2 and E36.12.4 cells were incubated for 1 h at 41 °C or 37 °C, metabolically pulse-radiolabelled and chased at 37 °C. H-2K^b was immunoprecipitated with Y-3 and analysed by two-dimensional IEF/SDS-PAGE. Immunoprecipitates made at the 0 time point of the chase were analysed and showed a single immature heavy-chain spot which chased into the more acidic and larger molecular mass forms (data not shown). H-2K^b heavy and light chains are indicated as above.

METHODS. *a*, 4×10^7 cells were incubated for 1 h at 41 °C or 37 °C in methionine-free media and then radiolabelled for 15 min at 37 °C with 0.7 mCi of [³⁵S]methionine ($>1,000$ Ci mmol⁻¹, NEN). Detergent lysis, immunoprecipitations and SDS-PAGE were done as described²², except that additional preclearing steps with IgG-sorb (5 times) and PA-Sepharose (1 time) were included before and after the first immunoprecipitation, respectively. *b*, Cells were incubated and pulse-radiolabelled for 10 min with [³⁵S]methionine as described above. The cells were then chased in the presence of cold methionine (15 mg ml⁻¹) for 50 min at 37 °C. Cell lysates were prepared and immunoprecipitated with Y-3. The immunoprecipitates were analysed by two-dimensional IEF/SDS-PAGE, using the Immobiline DryStrip Kit (Pharmacia). Gels were incubated in Autofluor (National Diagnostics), and radiolabelled proteins visualized by autoradiography at -80 °C.

of free heavy chains (Fig. 2*a*) and β_2 -microglobulin (data not shown) that were potentially available for assembly with peptide were also found at both temperatures.

The transport of assembled K^b molecules was analysed by pulse-radiolabelling and chasing at 37 °C. Elevated temperature did not affect the degree of charge or molecular mass heterogeneity of K^b molecules in either mutant or wild-type cells (Fig. 2*b*) as determined by two-dimensional isoelectric focusing/sodium dodecyl sulphate-polyacrylamide gel electrophoresis (IEF/SDS-PAGE). Because changes in charge and molecular mass during the intracellular maturation of class I molecules are associated with complex sugar additions in the Golgi complex, these findings together with those above indicated that synthesis, assembly and transport of H-2K^b occurred normally in mutant cells after incubation at 41 °C.

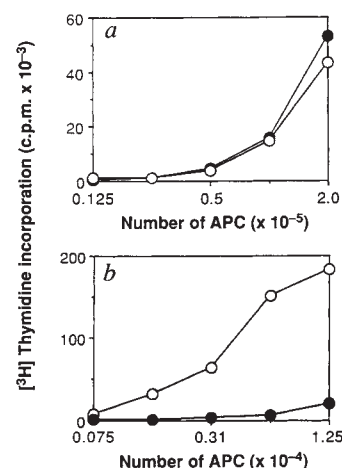
We next infected the mutant cells with a recombinant vaccinia virus containing a minigene encoding the OVA₂₅₇₋₂₆₄ peptide (OVA₂₅₇₋₂₆₄Vac) preceded by an initiating methionine and assayed antigen presentation. Presentation of the endogenously synthesized OVA₂₅₇₋₂₆₄ peptide was unaffected by incubation at

41 °C (Fig. 3*a*), whereas, as shown earlier, the presentation of ovalbumin was inhibited (Fig. 3*b*). Overproduction of the endogenously synthesized OVA₂₅₇₋₂₆₄ peptide is an unlikely explanation for the lack of effect of the 41 °C incubation because presentation by the vaccinia-infected mutant cells was considerably less than the presentation by cells that were incubated with ovalbumin (compare the 37 °C groups in Fig. 3*a* and *b*, and see the legend to Fig. 3). This finding indicates that incubating mutant cells at 41 °C did not affect transport of OVA₂₅₇₋₂₆₄ peptide from the cytosol into the endoplasmic reticulum, assembly of the OVA₂₅₇₋₂₆₄ peptide and H-2K^b in the endoplasmic reticulum, or transport of the class I/peptide complex to the cell surface.

Our results indicate that ubiquitin-conjugation plays an important role in the presentation of a cytosolic antigen with major histocompatibility complex (MHC) class I molecules. The essential role of ubiquitin conjugation in the degradation of many cytosolic proteins has led to the suggestion that it may be involved in class I antigen processing⁵. In fact, an amino-terminal modification of a viral protein that promotes ubiquitin-

FIG. 3 MHC class I-restricted presentation of endogenously synthesized OVA₂₅₇₋₂₆₄ peptide. Ts20.10.2 cells were incubated at 41 °C (filled circles) or 37 °C (open circles) for 1 h and then *a*, infected with OVA₂₅₇₋₂₆₄Vac for 1 h at 37 °C or *b*, incubated with OVA as described in the legend to Fig. 1. The indicated number of APCs were incubated with RF33.70.

METHODS. *a*, Ts20.10.2 cells (7×10^5 cells per ml) were incubated for 1 h at 41 °C or 37 °C and then infected with recombinant vaccinia virus (10 PFU per cell) for 1 h at 37 °C. *b*, Ts20.10.2 cells were incubated at 41 °C or 37 °C, treated with OVA, and then incubated at 37 °C as described in the legend to Fig. 1. APCs infected with a recombinant vaccinia virus containing an influenza nucleoprotein gene did not stimulate RF33.70. The thermolability of OVA-class I antigen presentation by ts20.10.2 cells was not altered by infection with the influenza-vaccinia recombinant virus (data not shown). Time-course experiments with OVA₂₅₇₋₂₆₄Vac showed that the amount of OVA₂₅₇₋₂₆₄ peptide-K^b complexes on the surface of cells infected for 1 h was limiting because cells infected for 3 h were 64-fold more efficient at stimulating RF33.70. OVA₂₅₇₋₂₆₄Vac constructed by inserting a synthetic oligonucleotide encoding the peptide MSIINFEKL behind the vaccinia virus p7.5 early-late promoter in pSc11 (refs 36 and 37) was provided by R. Anderson, J. Yewdell and J. Bennink.



dependent degradation enhanced its presentation with class I¹³. The inhibition in presentation of cytosolic ovalbumin but not cytosolic endogenously synthesized OVA peptide in ts20.10.2 cells exposed to 41 °C implies that the thermolabile block in these cells is in the production of antigenic peptide from the intact protein. The simplest conclusion from the present findings is that ubiquitin conjugation to ovalbumin plays a critical role in the generation of peptides for antigen presentation. But it is possible that ubiquitin, either directly or indirectly, is required for some other early step in the presentation of this antigen.

The assembly of stable class I molecules requires the binding of peptides to class I heavy chains and β_2 -microglobulin^{22,24}. Our finding that the assembly of class I molecules in mutant cells was unaffected by the non-permissive temperature may indicate that: (1) the ubiquitin-dependent pathway was still producing other peptides, because ubiquitin conjugation was not completely blocked under our conditions (data not shown); (2) a ubiquitin-independent pathway²⁵, perhaps also using the proteasome⁴, was generating cytosolic peptides; and/or (3) protein antigens previously conjugated to ubiquitin or leader sequences in the endoplasmic reticulum²⁶ were still being processed. Alternatively, class I molecules lacking peptides can form stable complexes with xenogeneic β_2 -microglobulin²⁷. Consistent with this latter possibility is the finding that K^b on these transfected hamster cells is at least 500-fold more efficient in presenting exogenous peptides than mouse cells in the absence of β_2 -microglobulin²¹ or cellular metabolism²¹ (data not shown). Our data do not resolve whether ubiquitin-dependent proteolysis is also critical in processing other antigens or whether there is a difference in the processing of biosynthesized antigens versus exogenous antigens introduced into the cytosol. Analogous experiments with ts20.10.2 cells should be useful in further defining the role of ubiquitin in antigen processing.

The presence of two MHC-encoded proteins in the proteasome has led to the hypothesis that the proteasome is involved in processing cellular and viral antigens^{14,15}. Recent data, however, indicate that these subunits are not essential for class

I-restricted antigen presentation^{28,29}. Because ubiquitin-conjugated proteins are degraded by the 26S proteasome complex^{3,4}, our findings directly implicate this structure in processing of an antigen for class I-restricted presentation. □

Received 27 October 1992; accepted 29 March 1993.

- Hershko, A. & Ciechanover, A. A. *Rev. Biochem.* **61**, 761–807 (1992).
- Rechsteiner, M. A. *Rev. Cell Biol.* **3**, 1–30 (1987).
- Goldberg, A. L. *Eur. J. Biochem.* **203**, 9–23 (1992).
- Goldberg, A. L. & Rock, K. L. *Nature* **357**, 375–379 (1992).
- Townsend, A. R. M., Gotoh, F. M. & Davey, J. *Cell* **42**, 457–467 (1985).
- Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324**, 575–577 (1986).
- Morrison, L. A. *et al. J. exp. Med.* **163**, 903–921 (1986).
- Moore, M. W., Carbone, F. R. & Bevan, M. J. *Cell* **54**, 777–785 (1988).
- Spies, T. & Demars, R. *Nature* **351**, 323–324 (1991).
- Powis, S. J. *et al. Nature* **354**, 529–531 (1991).
- Yewdell, J. W. & Bennink, J. R. *Adv. Immun.* **52**, 1–123 (1992).
- Monaco, J. J. *Immun. Today* **13**, 173–179 (1992).
- Townsend, A. *et al. J. exp. Med.* **168**, 1211–1224 (1988).
- Parham, P. *Nature* **348**, 674–675 (1990).
- Yang, Y., Waters, J. B., Fruh, K. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4928–4932 (1992).
- Kulka, R. G. *et al. J. Biol. Chem.* **263**, 15726–15731 (1988).
- Gropper, R. *et al. J. Biol. Chem.* **266**, 3602–3610 (1991).
- Rock, K. L., Rothstein, L. & Gamble, S. J. *Immun.* **145**, 804–811 (1990).
- Wiley, H. S. & McKinley, D. N. *Meth. Enzym.* **146**, 402–417 (1987).
- Falk, K. *et al. Nature* **351**, 290–296 (1991).
- Rock, K. L., Rothstein, L. & Benacerraf, B. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8918–8922 (1992).
- Townsend, A. *et al. Cell* **62**, 285–295 (1990).
- Smith, M. H. & Barber, B. H. *Molec. Immun.* **27**, 169–180 (1990).
- Ljunggren, H.-G. *et al. Nature* **346**, 476–480 (1990).
- Van Pel, A. & Boon, T. *Immunogenetics* **29**, 75–79 (1989).
- Henderson, R. A. *et al. Science* **255**, 1264–1266 (1992).
- Alexander, J. *et al. Immunogenetics* **31**, 169–178 (1990).
- Arnold, D. *et al. Nature* **360**, 171–173 (1992).
- Momburg, F. *et al. Nature* **360**, 174–177 (1992).
- Michalek, M. T., Benacerraf, B. & Rock, K. L. *J. Immun.* **146**, 449–456 (1991).
- Takebe, Y. *et al. Molec. cell. Biol.* **8**, 466–472 (1988).
- Siu, G., Hedrick, S. M. M. & Brian, A. A. *J. Immun.* **143**, 3813–3820 (1989).
- Dang, L. H. *et al. J. Immun.* **144**, 4082–4091 (1990).
- Kohler, G., Fischer-Lindahl, K. & Heusser, C. *Immun. System* **2**, 202–208 (1981).
- Jones, B. & Janeway, C. A. Jr *Nature* **292**, 547–549 (1981).
- Chakrabarti, S., Brechling, K. & Moss, B. *Molec. cell. Biol.* **5**, 3403–3409 (1985).
- Moss, B. & Eard, P. in *Current Protocols in Molecular Biology* (eds Ausubel, F. M. *et al.*) 16.15.1–16.19.9 (Greene-Wiley, New York, 1991).

ACKNOWLEDGEMENTS. We thank R. Kulka for ts20 and E36 cell lines, B. Barber for anti-exon 8 antiserum, R. Anderson, J. Yewdell and J. Bennink for the gift of OVA_{257–264} Vac, and M. J. Tawa for help in preparing the manuscript.

Release of secretory products during transient vesicle fusion

G. Alvarez de Toledo*, R. Fernández-Chacón* & J. M. Fernández†

* Departamento de Fisiología Médica y Biofísica, Facultad de Medicina, Universidad de Sevilla, Avda. Sánchez Pizjuán, 4 41009 Sevilla, Spain

† Department of Physiology & Biophysics, Mayo Clinic, Rochester, Minnesota 55905, USA

PATCH-CLAMP experiments have shown that fusion of secretory granules with the plasma membrane does not always occur as an all-or-none event, but can develop slowly in a fluctuating manner or can be transient^{1–5}. These observations suggested that release could be detected during such incomplete fusion events. To test this hypothesis we have combined patch-clamp measurements of the activity of single exocytotic fusion pores in beige mouse mast cells with the electrochemical detection of serotonin released during the exocytotic events. We report here that on fusion pore opening there is a small release of serotonin which is directly proportional to the pore conductance. We also show that a significant release occurs during transient fusion events. These results demonstrate, to our knowledge for the first time, release of a neurotransmitter from a secretory vesicle that did not undergo complete fusion.

Vesicle fusion and the fusion pore conductance were measured by following the admittance of a beige mouse mast cell in the whole-cell mode of the patch-clamp technique^{6,7}. Release of secretory products was measured with amperometry

by placing a carbon fibre electrode very close to the cell surface^{8–11}. Figure 1a (top) shows the changes in the cell membrane capacitance on fusion of single secretory vesicles with the plasma membrane. Three types of exocytotic events represented the three modes of fusion pore activities. First, a transient capacitance increase (Fig. 1a, arrow) that resulted from the opening and subsequent closure of the fusion pore. During this event, a small but significant amperometric signal was observed (Fig. 1a, amplified in b). Second, an irreversible capacitance increase that developed slowly (Fig. 1a, circle). These events resulted from the slow and fluctuating expansion of the fusion pore after its formation. A large, spike-like, amperometric signal was observed at the end of this event when the pore was fully open and the capacitance signal had reached its maximum value. But during the pore expansion phase, while the pore was still small and fluctuating (flicker), a small but measurable amperometric foot was observed. This foot was typically only a small fraction of the 'spike' (Fig. 1a, amplified in c). The third type of event was represented by a step increase in capacitance representing a fusion pore that instantly expanded fully. Accompanying such events we always observed a large amperometric spike with a fast rising phase and no foot preceding it. A comparison between fast cyclic voltammograms obtained either by direct application of serotonin to the carbon electrode (Fig. 1d, top plot), or during one of the spike-like exocytotic transients (Fig. 1d, bottom plot) suggests that serotonin is the main secretory product detected electrochemically during beige mouse mast cell secretion.

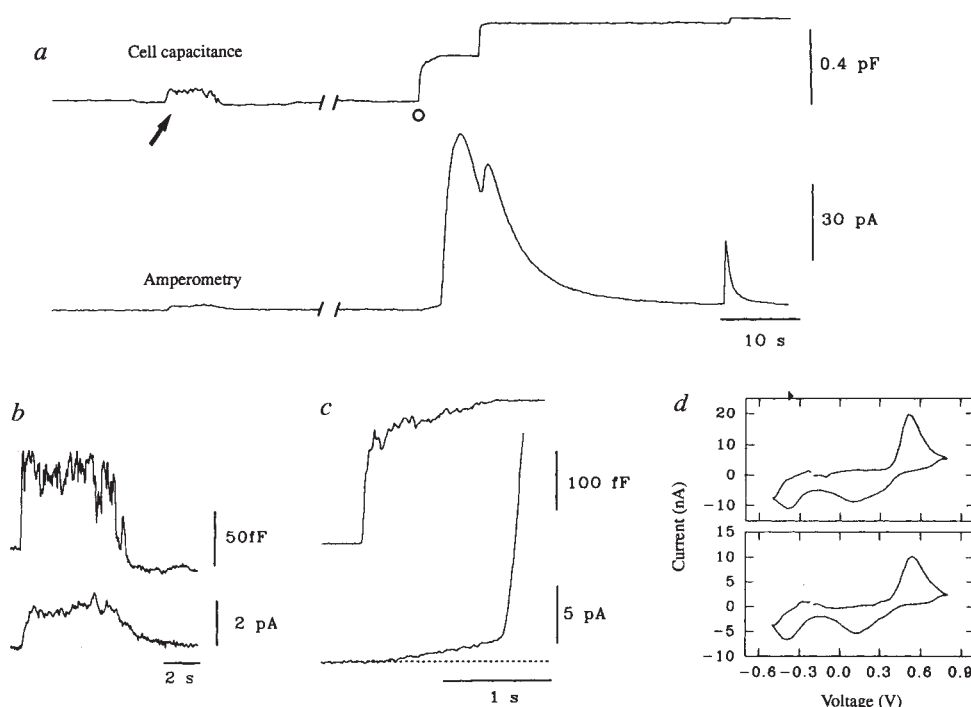
Figure 2 shows a detailed analysis of the fusion pore conductance and the amperometric signal from a vesicle that fused transiently before undergoing irreversible fusion. The vesicle

flickered for half a second (Fig. 2a, first trace) and during this period a significant amperometric signal was measured (Fig. 2a, second trace, amplified in *b*, fourth trace). The second trace of Fig. 2b represents the real part of the admittance from which the conductance of the pore, G_p , was calculated (Fig. 2b, third trace), assuming an electrical equivalent circuit for a fusing vesicle as illustrated in Fig. 2c. During flicker the conductance of the pore was continually changing between 0.21 and 1.5 nS. The amperometric signal closely followed the conductance changes, indicating that the size of the pore is limiting for release. Figure 2d shows that there is a linear relationship between the amplitude of the amperometric signal and the pore conductance and that secretion occurred even through a very narrow pore (close to its initial value of 230 pS)². The amount of secretory products diffusing out of the vesicle during flicker represented

only a small fraction of the material stored inside the vesicle because the integrated amperometric current during flicker is only 1% of that during the spike. This value is surprisingly low because if all serotonin molecules detected with amperometry (foot plus spike) were free to diffuse through the fusion pore on its formation, one would expect to see a linear relationship between pore conductance and release as shown by the dotted line of Fig. 2d. The experimental data show a reduced slope, suggesting that only a fraction of the serotonin contained in the secretory vesicle is available for release during flicker. Although other mechanisms, such as restricted diffusion of serotonin molecules through a pore of limited size, could explain this phenomenon, this observation fits well with the hypothesis that proposes that most of the secretory material, before exocytosis, is bound to the proteoglycan granule matrix and therefore is

FIG. 1 Simultaneous measurements of vesicle fusion and serotonin release by single secretory vesicles. *a*, Vesicle fusion appears as stepwise increases in the cell membrane capacitance (top trace); the serotonin released from the same vesicle is measured by the amperometric signal (bottom trace). A transient vesicle fusion is shown in *a* (top trace, arrow). An expansion of this recording is shown in *b*. In 9 out of 19 transient fusion events (18 different cells) a significant amperometric signal was detected. The average number of molecules detected during these events was $7 \pm 5 \times 10^6$ ($n=9$, mean $\pm$ s.e.m., assuming one charge per molecule). In 10 out of 19 cases we observed transient fusion events that did not cause an observable serotonin release. Such events may have occurred too far from the carbon electrode for their detection, or could correspond to granules with a much lower concentration of free serotonin. We cannot distinguish among these possibilities. A total of 148 irreversible fusion events from 21 different cells were observed as step increases in the cell membrane capacitance. In these events an amperometric spike was observed. In 19 events we observed a clear 'foot' preceding the main amperometric spike. These events showed a characteristically slow expansion of the fusion pore that coincided with the foot. The foot had a mean amplitude at its maximum value of 5.3 ± 4.9 pA. The amperometric spike had a peak amplitude of 182.4 ± 21.6 pA, a rising phase of 1.04 ± 0.2 pA ms⁻¹ and half-width (width of the peak at half its maximum amplitude) of 0.48 ± 0.12 s (mean $\pm$ s.e.m., $n=19$). The amperometric spike of the events that did not reveal a foot had a smaller amplitude, slower rising and decaying phases, and longer half-widths. They were characterized by a delay, as measured between the onset of capacitance and amperometric signals, of 206 ± 40 ms, a peak amplitude of the amperometric transient of 61 ± 5.5 pA, rising phase of 0.36 ± 0.045 pA ms⁻¹ and a half-width of 1 ± 0.1 s (mean $\pm$ s.e.m., $n=129$). The top graph in *d* shows a fast cyclic voltammogram obtained by direct application of serotonin (10μ M), dissolved in the external recording solution, onto the carbon fibre. This voltammogram typically revealed one oxidation (around 450 mV) and two re-reduction peaks. A similar voltammogram is obtained during beige mast cell exocytosis (bottom graph). The voltammogram was obtained by applying voltage ramps ranging from -0.5 to $+0.8$ V at a rate of 170 V s⁻¹. Other substances known to be present in mast cell granules (histamine and heparin), tested at concentrations of 1 mM, gave no voltammetric response. The break in the recording shown in *a* corresponds to 94 s.

METHODS. Beige mouse mast cells (bgj/bgj strain, The Jackson Laboratory, Bar Harbor, ME) were obtained by peritoneal lavage and incubated at 37 °C. In most experiments, the cells were incubated with 100μ M serotonin to enhance their vesicular contents because mast cells pump amines into their secretory vesicles¹⁹. For electrical recordings the cells were dialysed through a patch pipette with a solution containing (in mM): 140 K-glutamate, 7 MgCl₂,



10 HEPES, 3 KOH, 0.2 Mg-ATP, 10 Ca-EGTA buffers. The concentration of intracellular calcium was 35 nM. Exocytosis was triggered by adding 5μ M GTP- γ S to the pipette solution. The external recording solution was composed of (in mM): 140 NaCl, 10 HEPES, 3 KOH, 2 MgCl₂, 1 CaCl₂ and 10 glucose. The pH was 7.2 and the osmolality $300\text{--}310$ mmol kg⁻¹. All experiments were done at room temperature ($20\text{--}27$ °C). For amperometric measurements, glass-sealed carbon electrodes were prepared from $8\text{-}\mu$ m diameter carbon fibres following a procedure described previously²⁰. The fibre was held at a voltage of $+650$ mV. At this voltage, serotonin is the main substance, contained in beige mouse mast cell vesicles, capable of being oxidized¹¹; therefore, the amperometric signal mostly represents serotonin release. The redox current was measured with an EPC-7 patch-clamp amplifier. The current output of the amplifier was directly fed into a 15-bit A/D converter and filtered at 20 Hz. To monitor the cell membrane capacitance we used pipettes coated with a silicon elastomer. The cell membrane capacitance was measured with a software-based phase detector²¹. Briefly, a continuous sinusoidal voltage (800 Hz, 54 mV peak to peak), superimposed to a holding potential of 0 mV, was applied to the cell through the V-command input of a patch-clamp amplifier. The resulting current was then filtered at 2 kHz and sampled by a 15-bit A/D converter. Lock-in amplifier operations were done digitally. The phase was set such that one of the outputs of the lock-in reflected changes in membrane conductance and access resistance, while the other output reflected changes in the cell membrane capacitance. One data point was obtained by integration of eight sinusoidal cycles (one capacitance point every 9.5 ms). Lock-in operation and amperometric measurements were combined into the same program.

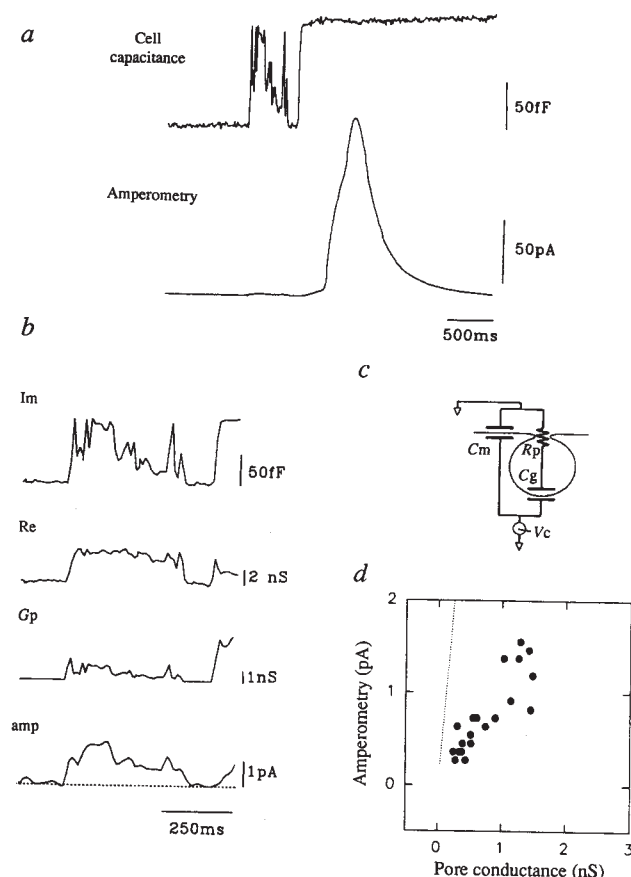


FIG. 3 Release of serotonin during slow expansion of a fusion pore before completing irreversible fusion. *a*, During the early phase of fusion, when the pore is still small and expanding, a small release of serotonin is seen to precede the large amperometric spike that marks the release of the total contents of the vesicle. *b*, Similar to Fig. 2, the real (Re) and imaginary (Im) component of the changes in cell admittance are used to calculate the conductance of the fusion pore (G_p) during these early moments of expansion. The fusion pore conductance, G_p , is shown to be well correlated with the amperometric current (amp). *c*, The filled circles show that there is a linear relationship between G_p and the release of serotonin in the experiment shown in *b*. A total concentration of free serotonin of 160 mM is calculated for this vesicle. If this serotonin was free to diffuse through the fusion pore, a linear relationship was predicted (dotted line) with a much steeper slope than that observed experimentally. Three other fusion events (open squares and circles, filled triangles) of similar characteristics are shown for comparison. In this figure each value is the average of six consecutive data points from the original recordings.

FIG. 2 Release of serotonin during transient fusion events. *a*, Vesicle fused transiently, then reopened (flicker) and finally fused irreversibly with the plasma membrane. During these events serotonin release was detected by the amperometric measurements. *b*, Real (Re) and imaginary (Im) components of the change in cell admittance, measured during the transient fusion event shown in *a*, are used to calculate the fusion pore conductance (G_p). The fusion pore conductance, G_p , is well correlated with the amperometric current (amp) which measures the serotonin released during this event. The fusion pore conductance, G_p , was calculated as $G_p = (Im/Re)\omega C_g$, where G_p is the pore conductance in nS, Im and Re are the absolute magnitude of the imaginary and real components of the admittance, ω is constant and has a value of $6,283 \text{ rad s}^{-1}$ and C_g is the capacitance of the granule in fF. The computation of G_p assumes that a fusion event is well represented, electrically, by the equivalent circuit shown in *c* where C_g is the capacitance of the vesicle, R_p , the pore resistance, C_m , the cell membrane capacitance, and V_c the membrane potential. There is a linear relation between pore conductance and release of serotonin (*d*, filled circles). All transient fusion events analysed ($n=6$) show a linear relation albeit with different slopes. From the total charge detected with the amperometric signal (large spike plus leak, 75.4 pC), the size of the vesicle (133 fF) and assuming the vesicle is a perfect sphere ($4.58 \mu\text{m}^3$), we calculate a minimal concentration of 170.6 mM inside the vesicle before exocytosis. The flux (i) of molecules through a pore was calculated from the following equation: $i = ((\pi P^2 D)/(Z + \pi P/2))(C_1 - C_0)$, where Z is the length of the pore (assumed 15 nm for two lipid bilayers), P , is the pore radius (calculated from G_p), D is the diffusion coefficient ($0.95 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ for a molecule of around M_r 200) and $C_1 - C_0$ is the concentration gradient across the pore. C_0 was assumed 0. Using this equation and a concentration gradient of 170.6 mM, we predicted the relationship between the pore conductance and serotonin flux to be linearly related (dotted line in *d*) with a slope much larger than that measured experimentally.

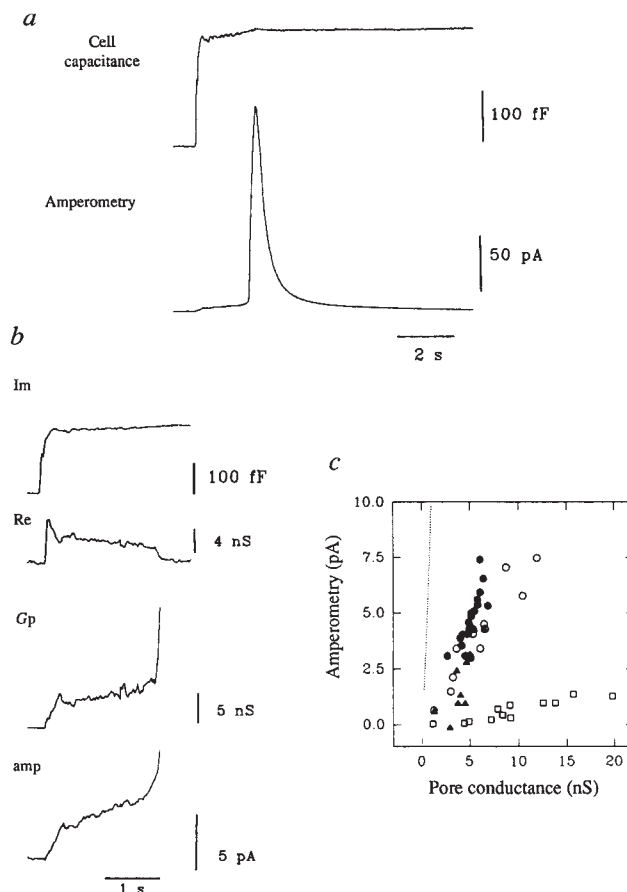
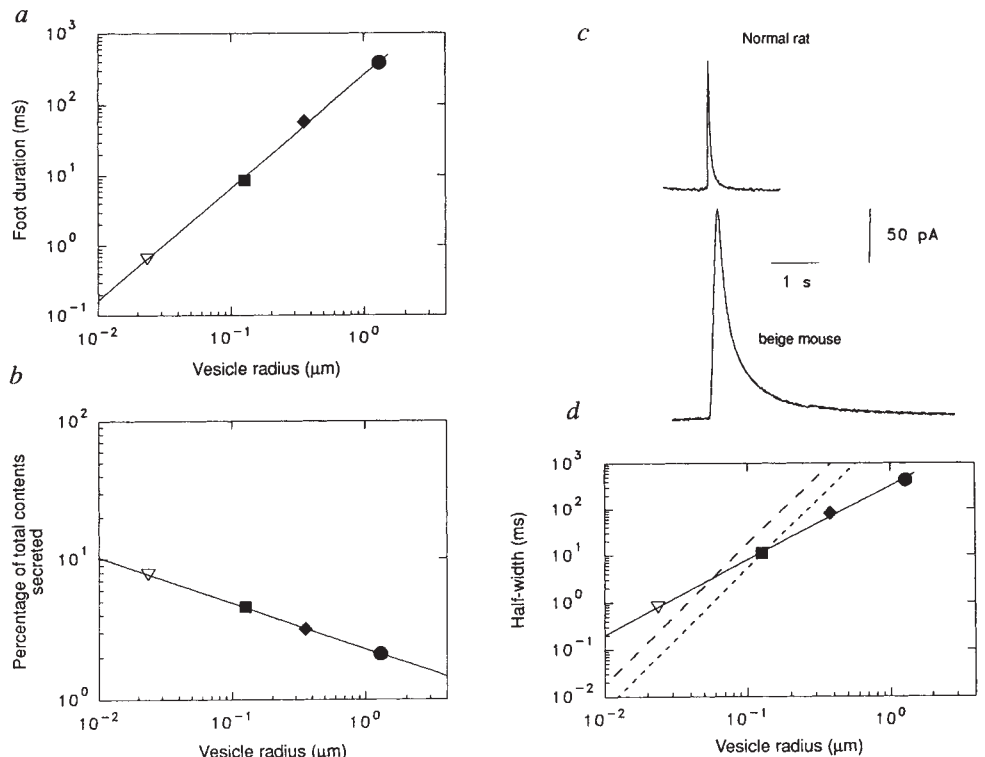


FIG. 4 The time course of release through the exocytotic fusion pore depends on the size of the secretory vesicle. *a*, Duration of the foot for three types of vesicles (beige mouse mast cells, filled circle; normal rat mast cells, filled diamond and chromaffin cells, filled square) are shown as a function of vesicle radius. Open triangles are predictions for a synaptic vesicle. The mean lag period between the capacitance change and the onset of the amperometric spike for beige mouse was 367 ± 64 ms ($n=148$). In 12.8% of these events a foot was distinguished with a mean duration of 410 ± 12 ms ($n=19$). In normal rat mast cells vesicles the mean latency between the capacitance change and the amperometric transient was of 58 ± 15 ms ($n=39$). Out of a total of 110 amperometric transients observed only 5 (4.5%) showed a foot with a mean duration of 66 ± 3 ms. The data from chromaffin vesicles was obtained from ref. 10. The predicted foot duration for a synaptic vesicle would be 0.55 ms (open triangle). *b*, Percentage of the total contents released during the foot as a function of vesicle radius (2.1 ± 1.6 , $n=19$, for beige vesicles, 3.2 ± 2 , $n=5$, for normal rat vesicles, 4.5 for chromaffin cell vesicles¹⁰). This percentage was obtained by dividing the foot charge by the charge of the total amperometric signal (foot plus spike). Symbols are as in *a*. An extrapolation to the size of a synaptic vesicle predicts that only 8% of the contents would be released during this phase (open triangle). The time course of release from the granule matrix, represented by the amperometric spike, also depends on the size of the vesicle. *c* (top), An example of a transient from a normal rat mast cell vesicle, the average half-width is 80 ± 3 ms ($n=110$,



average radius of 0.35 μm). A beige amperometric transient (*c*, bottom trace) is slower (480 ± 120 ms, $n=19$). The correlation between half-widths and vesicle radius is shown in *d*. If all molecules were in solution inside the vesicle the expected half-width of the amperometric transient would be as indicated by the dashed lines (medium dash, pore conductance of 300 pS; short dash of 1 nS). These calculations were done following equation (4) of ref. 17.

osmotically inactive and unavailable for release. The release of secretory products is thought to be associated with swelling of the matrix¹²⁻¹⁴. Because no measurable swelling of the granule matrix occurs during transient fusion events⁴, it is likely that the serotonin released during these events is not bound to the matrix and is dissolved in the small fraction of water existing inside the vesicle. The large spike of serotonin released would then correspond to the release from the swollen matrix.

Figure 3*a* shows an example of an amperometric signal detected during an irreversible fusion event where the foot lasted for 2 s. The time course of the pore conductance (Fig. 3*b*, third trace), obtained from the imaginary and real components of the admittance, shows that during the time that a foot was detected, there was a continuous widening of the exocytotic fusion pore. The amperometric signal during this foot was also well correlated with the conductance of the pore (Fig. 3*c*, filled circles). As with the transient fusion events, secretion through the pore seems to be limited by other factors than the pore conductance itself, as is suggested by the slope of the data points compared to the one expected if all serotonin molecules were in solution (dotted line, Fig. 3*c*). When other events were analysed (filled triangles, open circles, open squares, Fig. 3*c*), we found that the slope of the amperometric current/pore conductance relationship was different for each vesicle, suggesting that the intravesicle concentration of free serotonin varied among vesicles. The leak of secretory products detected during the expansion of the mast cell exocytotic fusion pores agrees with the foot observed in adrenal chromaffin cells, therefore confirming the hypothesis of a leak of substances through a narrow fusion pore¹⁰.

Our results show that release of secretory products from beige

mast cell granules depends on the activity of the fusion pore and occurs in two steps. First, a slow release of secretory material that is directly related with the conductance (radius) of the pore connecting the vesicle lumen with the exterior. This secretion accounts for the serotonin release measured immediately after the pore is formed, and is not accompanied by appreciable changes in the diameter of the granule matrix. We favour the interpretation that this release occurs from a pool of molecules dissolved in the fraction of water existing inside a mature vesicle. The second stage of secretion occurs when the granule matrix swells and the serotonin bound to the matrix can freely diffuse to the external medium. This release produces the fast and large amperometric transient (spike).

The percentage of the contents secreted by a vesicle during the expansion of the fusion pore depends on the size of the vesicle from which it is released and the time taken by the pore to expand (Fig. 4). For example, in beige mouse mast cell vesicles the foot lasts on average 410 ms (Fig. 4*a*, filled circle), secreting only 2.1% of their contents (Fig. 4*b*, filled circle). Experiments done in normal rat mast cells reveal that the foot lasts on average 66 ms, releasing 3.2% of the vesicle's contents (Fig. 4*a*, *b*, filled diamonds). Data taken from chromaffin cells¹⁰ fit well with the trend of the mast cell vesicles (Fig. 4*a*, *b*, filled squares), suggesting that foot duration is inversely related to the radius of the vesicle. If the mechanisms governing the leak through the pore were the same for a synaptic vesicle, the mean foot duration would be of 0.55 ms for a vesicle of 24 nm radius and the expected release would be 8% of the total contents (Fig. 4*a*, *b*, open triangle). These predictions are based on secretory vesicles which contain a dense core gel that avidly traps secretory products^{15,16}. This interaction may change the time course of

release. For example, a diffusional mechanism of release predicts that on the opening of a fusion pore, the flux of soluble granule components will decline exponentially, with a time constant proportional to the volume of the vesicle. Beige mouse mast cell granules have a volume that is on average 50 times larger than the wild-type granules ($10\ \mu\text{m}^3$ versus $0.2\ \mu\text{m}^3$). But in beige, the half-width of the amperometric spike is only five times larger than those observed in normal cells (Fig. 4c). In contrast, if the amperometric spike corresponded to the emptying of the vesicle through similar fusion pores (a reasonable assumption in both types of mast cells) a 50-fold difference in the half-width is predicted. Our data pooled together with similar data obtained from chromaffin cells¹⁰ are well described by a time constant of release proportional to the 1.5 power of the radius (Fig. 4d, solid line), in contrast to the third-power relationship predicted by simple diffusion (dashed lines). A release mechanism that is dominated by swelling of a gel is advantageous for large vesicles where simple diffusion would take a prohibitively long time. In contrast, a small vesicle would release material very fast only if its contents are in solution, being able to release its quantal contents during a brief opening. In such a situation, a vesicle could discharge its contents with a time constant of 0.24 ms, through a pore of 300 pS (ref. 17). This idea is consistent with the appearance of clear vesicles (apparently without a core) in fast synapses, as occurs at the neuromuscular junction¹⁸. The release of material through a pore without ever undergoing complete fusion would be advantageous for the rapid and selective retrieval of vesicle membrane. By contrast, synaptic vesicles with a dense core may behave more like chromaffin or mast cell

granules, releasing only a small fraction of their contents during transient fusion events or during the expansion phase of an irreversible event. □

Received 6 January; accepted 1 April 1993.

1. Fernández, J. M., Neher, E. & Gomperts, B. D. *Nature* **312**, 453–455 (1984).
2. Breckenridge, L. & Almers, W. *Nature* **328**, 814–817 (1987).
3. Zimmerberg, J., Curran, M., Cohen, F. S. & Brodwick, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1585–1589 (1987).
4. Breckenridge, L. & Almers, W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1945–1949 (1987).
5. Alvarez de Toledo, G. & Fernández, J. M. *Cell Physiology of Blood* (eds Gunn, R. B. & Parker, J. C.) 333–344 (The Rockefeller University Press, New York, 1988).
6. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
7. Neher, E. & Marty, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6712–6716 (1982).
8. Wightman, R. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10754–10758 (1991).
9. Tatham, P. E. R., Duchon, M. & Millar, J. *Pflügers Arch.* **419**, 409–414 (1991).
10. Chow, R. H., von Rüden, L. & Neher, E. *Nature* **356**, 60–63 (1992).
11. Fernández-Chacón, R. & Alvarez de Toledo, G. *Biophys. J.* **61**, A226 (1992).
12. Fernández, J. M., Villalón, M. & Verdugo, P. *Biophys. J.* **59**, 1022–1027 (1991).
13. Curran, M. J. & Brodwick, M. S. *J. gen. Physiol.* **98**, 771–790 (1991).
14. Monck, J. R., Oberhauser, A. F., Alvarez de Toledo, G. & Fernandez, J. M. *Biophys. J.* **59**, 39–47 (1991).
15. Rabenstein, D. L. *Biochemistry* **26**, 6923–6926 (1987).
16. Daniels, A. J., Williams, R. P. J. & Wright, P. E. *Neuroscience* **3**, 573–585 (1978).
17. Almers, W., Breckenridge, L. J. & Spruce, A. E. *Secretion and its Control* (eds Oxford, G. & Armstrong, C.) 269–282 (The Rockefeller University Press, New York, 1989).
18. Heuser, J. E. & Reese, T. *Handbook of Physiology* Vol. 1 (ed. Kandel, E. R.) 261–294 (Williams & Wilkins, Bethesda, 1977).
19. Ludowyke, R. I. & Lagunoff, D. *Biochemistry* **25**, 6287–6293 (1986).
20. Millar, J. & Williams, G. V. *J. electroanal. Chem.* **282**, 33–49 (1990).
21. Joshi, C. & Fernández, J. M. *Biophys. J.* **53**, 885–892 (1988).

ACKNOWLEDGEMENTS. We thank J. Millar and M. Duchon for assistance in the manufacture of carbon electrodes and for providing carbon fibres. L. Tabares and P. Tatham for helpful comments on the manuscript, J. López Barneo for comments and loan of equipment and A. León Boza and C. Camrud for technical assistance. Supported by a DGICYT (Science and Technology Commission of Spain) grant to G.A.T. and NIH grants to J.M.F.

Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis

Yurij Ionov*, Miguel A. Peinado*†, Sergei Malkhosyan*, Darryl Shibata‡ & Manuel Perucho*§

* California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, California 92037, USA

‡ Department of Pathology, University of Southern California School of Medicine, Los Angeles, California 90033, USA

SPONTANEOUS errors in DNA replication have been suggested to play a significant role in neoplastic transformation and to explain the chromosomal alterations seen in cancer cells¹. A defective replication factor could increase the mutation rate in clonal variants arising during tumour progression, but despite intensive efforts, increases in tumour cell mutation rates have not been unambiguously shown². Here we use an unbiased genomic fingerprinting technique³ to show that 12 per cent of colorectal carcinomas carry somatic deletions in poly(dA · dT) sequences and other simple repeats. We estimate that cells from these tumours can carry more than 100,000 such mutations. Only tumours with affected poly(dA · dT) sequences carry mutations in the other simple repeats examined, and such mutations can be found in all neoplastic regions of multiple tumours from the same patient, including adenomas. Tumours with these mutations show distinctive genotypic and phenotypic features. We conclude that these mutations reflect a previously undescribed form of carcinogenesis in the colon (predisposition to which may be inherited) mediated by

a mutation in a DNA replication factor resulting in reduced fidelity for replication or repair (a 'mutator mutation').

The arbitrarily primed polymerase chain reaction (AP-PCR) is a PCR-based DNA fingerprinting technique using primers whose nucleotide sequence is arbitrarily chosen^{3,4}. Competition between the annealing events during the initial low-stringency cycles results in the reproducible and quantitative amplification of many discrete bands during the subsequent high-stringency cycles. The genomic fingerprint obtained in a simple experiment can identify somatic genetic alterations. AP-PCR is useful for the detection and isolation of tumour-specific allelic losses and gains, thus providing a molecular alternative to cancer cytogenetics⁵. Figure 1 shows the AP-PCR fingerprints obtained with several arbitrary primers of normal and tumour tissue DNAs from two colorectal cancer patients. The differences in band mobility shown by arrows were suggestive of somatic mutations because they were tumour-tissue-specific. This phenomenon was observed with 6 of 10 unrelated primers. Without exception, there was a reduction in band size in tumour tissue. Cloning and sequencing showed that bands with altered mobilities contained repeated Alu sequences, which had undergone deletions in their poly(A) tails (data not shown). These deletions also occurred in runs of dA · dT base pairs present in DNA sequences without Alu repeats. The 550-nucleotide (nt) band amplified with primers B and C (APΔ3) is an example (Fig. 2). Comparative sequencing and PCR analyses of APΔ3 genomic and cloned sequences after AP-PCR amplification, established that the sequences predominantly present in normal and tumour tissues of patient 197 contained 18 and 14 deoxyadenosines, respectively (data not shown). Therefore, a somatic deletion of 4 base pairs (bp) had occurred in APΔ3 sequences in these tumour cells.

Figure 3 shows the amplification by PCR of APΔ3 (b) and another sequence (APΔ2) containing an Alu repeat (a) from matched pairs of normal tumour DNAs. With the exception of cases 78 and 83, the bands corresponding to tumour tissue were a few nucleotides smaller than those from normal tissue. The detection of somatic deletions by AP-PCR implied that these mutations were very abundant, because of the random origin

† Present address: Institut de Recerca Oncologica, Hospital Duran i Reynals, Autovia de Castelldefels, Km 2.7 Hospitalet, 08907 Barcelona, Spain.

§ To whom correspondence should be addressed.

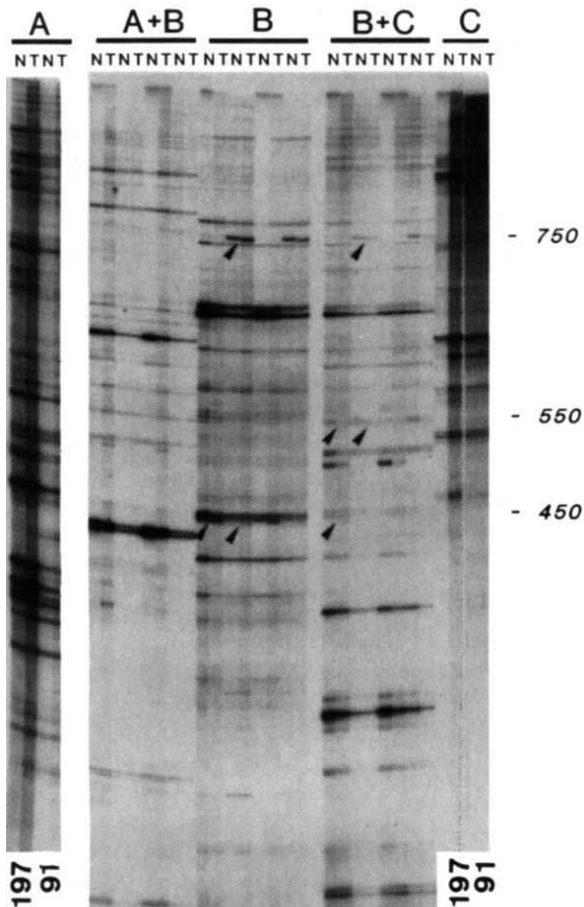


FIG. 1 AP-PCR analysis of colorectal carcinomas. Autoradiogram of the AP-PCR pattern of colorectal tumours 197 and 91. DNA from normal (N) and tumour (T) tissues were amplified with combinations of arbitrary primers A, B and C. For primers A+B, B and B+C, the samples were analysed in duplicate. Numbers at right indicate the rough molecular mass (in nucleotides, nt) of bands with small deletions in tumour DNA, indicated by arrowheads. METHODS. The origin of the tumour samples and the conditions for DNA preparation and amplification by AP-PCR have been described⁵. AP-PCR reactions were done with 125 μ M of each nucleotide, 2 μ M of arbitrary primers, 5 μ Ci [α -³²S]-dATP or 1 μ Ci [α -³²P]dCTP (New England Nuclear) and 2 units of *Taq* polymerase (Cetus/Perkin Elmer or Stratagene) in PCR buffer. The reaction consisted of 5 low-stringency cycles (1 min at 95 °C, 1 min at 50 °C, 1.5 min at 72 °C) and 30 high-stringency cycles (30 s at 60 °C, 1.5 min at 72 °C). The AP-PCR product was diluted 1:1 with formamide-loading buffer, denatured at 90 °C for 3 min and analysed on a 6% acrylamide, 8M urea sequencing gel. The arbitrary primers used in this study are listed below.

A (3U): 5'-AAGGGATCCCCCTTGCCGTC-3'
 B (K3US): 5'-GGAGTCGACTGGTGCTATACTTTT-3'
 C (6U): 5'-GTGAGCTCCTAGGTTGGCTCTGACT-3'
 J (ALU): 5'-GGTGAAATCCTGTCTCTACTAAAA-3'

of the sequences amplified from the cell genome due to the arbitrary nature of the primer sequences⁵. If one or two bands contained deletions with over half of the primers tested, these results implied that deletions occurred every $1-3 \times 10^4$ bp, because about 10-30 kilobases (kb) of the genome are probed with each primer. Therefore, the estimated number of sequences containing deletions would be $3 \times 10^9 / 1-3 \times 10^4 = 1-3 \times 10^5$.

This hypothesis predicted that deletions would occur in many other sequences containing runs of deoxyadenosines. We amplified by PCR the 3' region of the c-K-ras oncogene comprising an Alu repeat with a poly(A) tail of 12 deoxyadenosines⁶. Deletions were present in tumours with mutations in APΔ2 and APΔ3 sequences, whereas no deletions were found in the negative tumours (Fig. 3c), confirming the predictive value of the hypothesis. In addition to the amplification of multiple bands with deletions with the same arbitrary primer and the detection of deletions in both alleles of the same sequences, the ubiquitous presence of deletion mutations in these colorectal tumours was further supported by their detection in about one third of the genomic Alu repeats amplified by Alu-PCR⁷ (data not shown).

These frequent deletion mutations were not restricted to carcinomas, but also occurred in adenomas. Figure 3d shows deletions in APΔ2 sequences from two carcinomas and two adenomas, in all neoplastic areas of these specimens, but not in the immediately adjacent normal tissue (N). There were no obvious differences in the number of deleted nucleotides between the superficial (S) and the deeply invasive (D) areas of the carcinomas. These results strongly support the concept that these deletion mutations were somatic events that occurred before (or early after) neoplastic transformation and therefore that they were not the result of genetic instability of cancer cells during tumour progression.

Deletions in APΔ3 sequences were found in 16 of 137 (11.6%)

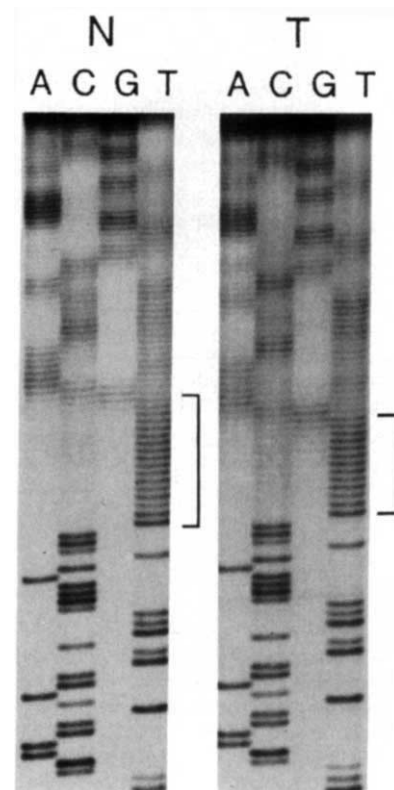


FIG. 2 Sequencing of APΔ3 from normal (N) and tumour (T) tissues from case 197. The genomic sequences were amplified by AP-PCR, the bands were excised from the gel, and reamplified with the same arbitrary primers B+C. The nucleotide sequence of APΔ3 DNA fragment excised from AP-PCR gels was determined with Sequenase as described⁵. Samples were diluted with formamide loading buffer, heat-denatured and aliquots of 2 μ l analysed in a sequencing gel for 3 h at 50W.

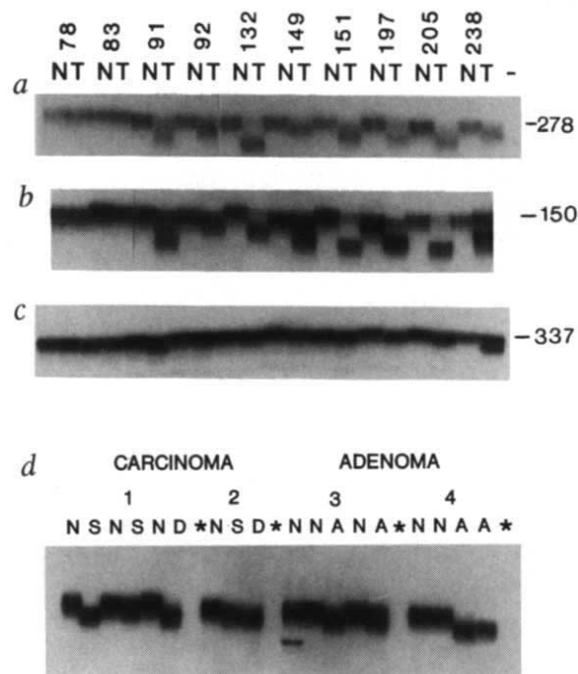


FIG. 3 PCR amplification of DNA sequences with somatic deletions. Autoradiogram of ^{32}P -labelled PCR products from AP Δ 2 (a), AP Δ 3 (b) and c-K-ras Alu (c) genomic DNA sequences from matched normal (N) and tumour (T) tissues from cases indicated at top. Numbers at right indicate the size (in nucleotides) of the most common alleles of these sequences. Deletions of AP Δ 2 sequences in synchronous carcinomas and adenomas (d). These four tumours appeared to be distinct entities based on their anatomical location and pathological examination. The samples were analysed in denaturing gels as in Fig. 1.

METHODS. The primers used for PCR amplification¹⁸ of the AP Δ 2 sequences were AP Δ 2U: 5'-TATCTGGGTCTGAATATGTCTTGA-3' and AP Δ 2D: 5'-TGATCATATTGAAATTTAAATCAAA-3'. The relative orientation of the primers (U, upstream; D, downstream) is given by the deoxyadenosine strand. This PCR generated a 460-bp DNA fragment comprising an Alu repeat. The J primer (see Fig. 1) internal to the Alu repeat, was used for a hemi-nested PCR amplification of a 278-bp DNA fragment after the initial PCR amplification (15 cycles) of the 460-bp fragment. This hemi-nested PCR was necessary to generate a DNA fragment sufficiently small to yield clear differences in mobility of bands differing in only a few nucleotides. The primers used for the PCR of the AP Δ 3 sequences were AP Δ 3U: 5'-GAGGCCAGCAATCT-GCACT-3' and AP Δ 3D: 5'-AAATCAGTATAAGAAAGGAA-3'. This PCR generated a DNA fragment of 150 bp, comprising the run of dA·dT. Amplification of the 3'-end untranslated sequences of the c-K-ras gene⁶ was done using PCR primers K3U: 5'-TAAATGAGTTCTGCAAAACAGG-3' and K3D: 5'-ATCTTC-ATGCAATGAAAAATAC-3'. PCR amplifications were done for 30 cycles (15 s at 94 °C; 15 s at 55 °C; 30 s at 72 °C) using [α - ^{32}P]-dCTP. AP Δ 2 sequences were also amplified (d) by PCR from small areas of the tumours (100–500 cells) protected from irradiation with ultraviolet of stained slides, by the method of ref. 19 as described²⁰. N, normal tissues; A, adenoma tissue; superficial (S) and deeply invasive carcinoma (D). Asterisks, PCR products from surrounding tissue not protected from ultraviolet irradiation.

carcinomas. We also analysed 78 and 40 of these tumours for deletions in AP Δ 2 and c-K-ras Alu sequences, respectively. In most positive tumours for AP Δ 3 deletions, there were also deletions in the other two sequences. Somatic mutations in these tumours were not exclusive of monotonic runs of dA·dT base pairs but also occurred in other di- (CT and CA) and trinucleotide (GGC) simple repeats (data not shown). Mutations in di- and trinucleotide repeats were restricted to the tumours with mutations in poly(dA·dT) sequences. A test for deviation of

the Poisson distribution indicated that these mutations were significantly clustered in the same tumours ($P < 0.0001$). We cannot rule out that these deletions could also occur with lower frequency in other tumours. However, based on accumulative data with various primers and with various repeated sequences, we estimate that the mutation incidence in these putative negative tumours must be ≥ 100 times lower than that of the positive tumours.

Although we have no clue as to the defect in the replication/re-

TABLE 1 Characteristics of colorectal carcinomas with ubiquitous somatic deletion mutations

Case																			Tumours with and without deletions		P
	43	51	61	68	73	89	91	92	132	149	151	197	201	205	211	238			+	–	
Age	56	67	55	73	72	39	53	37	71	80	78	57	54	49	63	70	(M/F)		60.87+/-12.9	65.62+/-10.2	0.049
Sex	M	F	F	M	F	M	F	M	M	F	F	M	F	M	M	M	(M/F)		9/7	79/42	0.32
Race	W	W	B	B	B	W	B	W	W	B	W	B	B	W	W	W	(W/B)		9/7	91/24	0.049
Differ.*	W	M	M	M	P		P	M	P	P	M	W	M	P	M	P	(P/WM)		6/9	17/90	0.036
Locat.†	C	A	A	T	C	A	T	R	C	C	C	T	C	R	C	A	(L/R)		2/14	71/38	0.00008
Invas.‡	A	A	B	B	B	A	B	B	B	B	D	C	B	B	B	A	(+/-)		2/14	68/61	0.002
Ras§	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	(+/-)		2/14	53/68	0.013
P53	–	–	–	–	–	–	N	+	–	+	–	–	–	+	–	+	(+/-)		4/11	62/45	0.022

All tissues were obtained from the Tissue Procurement Network of the University of Alabama as freshly frozen specimens collected from surgical resections with curative intent⁵. Case 89 is the only exception, obtained from New York University Medical College. The Case in Fig. 3d is not included because the various synchronous specimens were obtained as formalin-fixed paraffin blocks from the Pathology Department of the University of Southern California School of Medicine at Los Angeles.

* Degree of differentiation: W, well; M, moderate; P, poor. In mixed tumours, the presence of less differentiated phenotype was considered to be the dominant character.

† C, cecum; A, ascending; T, transversal; D, descending; S, sigmoid; R, rectum. L/R, left (D+S+R)/right (C+A+T).

‡ Turnbull's modification of Dukes' classification²¹. –/+; metastases absent (A+B)/metastases present (C+D). Includes 19 metastatic carcinomas to the liver, all negative for deletions.

§ Presence (+) or absence (–) of mutations at codons 12 and 13 of the c-K-ras or N-ras proto-oncogenes²².

|| Presence (+) or absence (–) of mutations in the p53 tumour suppressor gene²². All four tumours positive for p53 mutations retained the normal allele (4/0 versus 27/70; $P = 0.007$). Ubiquitous deletions correlated with cases of synchronous and metachronous tumours (4/12 versus 5/99, $P = 0.0018$).

Probabilities were calculated by the Fisher exact test in all cases but for the age, which was calculated by comparison of estimated pools of variance (T, –1,677; d.f. = 130).

N, Not analysed.

pair machinery of these tumour cells, these deletion mutations may be due to slippage during replication/repair by strand misalignment⁸. This model may explain the proportionality between the length of the runs and mutation frequency⁹ and the bias for deletions versus insertions¹⁰. We do not know if these microdeletions are exclusive to tumour cells. Simple repeated sequences are genetically unstable, as judged by their increased mutation rate *in vivo*⁹ and *in vitro*¹¹, and by their polymorphic nature in the human population¹². Because no deletions were found in most carcinomas, these mutations cannot be frequent events in the cells that originate most of the colorectal tumours. Therefore, the replication fidelity of these unstable sequences by the colon stem cells can be maintained in normal conditions for over 10,000 replications¹³.

The incidence of deletion mutations could be linearly related to the number of tumour cell replications. There was, however, no obvious correlation between the extent of the deletions, or the number of sequences with deletions, and the age of the positive cancer patients (data not shown). Together with the bimodal distribution of tumours with and without deletions, these are differences that distinguish our findings from the steady loss of telomere sequences during tumorigenesis and ageing^{14,15}.

We compared the characteristics of colorectal carcinomas with and without such mutations. The presence of generalized deletions correlated negatively with mutations in the p53 and c-K-ras genes and with tumours that had metastasized at diagnosis, and positively with poorly differentiated tumours, with tumours of the right colon and with tumours from blacks (Table 1). The two adenomas and two carcinomas of Fig. 3d were resected from the same patient, a 38-year-old black male. These results demonstrate that the widespread deletions are closely associated to neoplastic transformation.

We propose that the origin of these extensive microdeletions is in turn a mutation in a gene coding for a factor essential for the replication fidelity (or repair) of these simple repeated sequences. This mutation, occurring in a stem cell of the colon epithelium before (or soon after) neoplastic transformation, would lead to an increased mutation rate in the descendent cells. The negative correlation with p53 mutations, as well as the early occurrence of these deletions, rule out p53 as a possible obvious candidate for this mutator mutation. Therefore, it appears that the extensive genetic damage present in these tumours can escape from the proposed surveillance of genomic integrity by wild-type p53¹⁶.

The simultaneous presence in the same individual of at least four distinct tumours, all with generalized deletions, implies that the mutator mutation plays an active role in tumorigenesis. It also suggests that the occurrence of these mutations, although somatic, has an inherited predisposition. Therefore, it is possible that these ubiquitous deletions are the molecular genomic manifestations of a hereditary nonpolyposis colorectal cancer syndrome¹⁷. This could explain the early onset of some of these tumours and their striking prevalence in the proximal colon (Table 1). A property of the system that we have described is that once the initial mutational event occurs, the cell mutation rate may be so high that the premise for the rarity of clonal mutations in tumour cells no longer applies²³.

We have shown that the genetic instability of monotonic runs of dA·dT base pairs and other simple repeats is magnified in a subset of colon tumours. The distinctive characteristics of these tumours in various molecular, biological and clinical parameters, support the existence of a distinct mechanism for colorectal cancer development which involves the accumulation of these mutations, possibly as a result of a catastrophic loss of fidelity in the replication machinery of normal cells. The instability of the sequences undergoing mutations, the revealing marks of this loss of fidelity, could be exploited for the development of simple tests for the diagnostic detection, not only of tumour genetic instability, but also of its possible hereditary susceptibility. □

Received 15 February; accepted 5 April 1993.

- Loeb, L. A., Springgate, C. F. & Battula, N. *Cancer Res.* **34**, 2311–2321 (1974).
- Harris, C. C. *Cancer Res.* **51**, 5023–5044 (1991).
- Welsh, J. & McClelland, M. *Nucleic Acids Res.* **18**, 7213–7218 (1990).
- Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. *Nucleic Acids Res.* **18**, 6531–6535 (1990).
- Peinado, M. A., Malkhosyan, S., Velazquez, A. & Perucho, M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10065–10069 (1992).
- Kahn, S. M. *et al. Anticancer Res.* **7**, 639–652 (1987).
- Sinnett, D., Deregion, J.-M., Simard, L. R. & Labuda, D. *Genomics* **7**, 331–334 (1990).
- Streisinger, G. *et al. Cold Spring Harb. Symp. quant. Biol.* **31**, 77–84 (1966).
- Ripley, L. S. *A. Rev. Genet.* **24**, 189–213 (1990).
- Kunkel, T. & Soni, A. *J. biol. Chem.* **263**, 14784–14789 (1988).
- Kunkel, T. A. *J. biol. Chem.* **261**, 13581–13587 (1986).
- Orita, M., Sekiya, T. & Hayashi, K. *Genomics* **8**, 271–278 (1990).
- Armour, J. A. L., Patel, I., Thein, S. L., Fey, M. F. & Jeffreys, A. J. *Genomics* **4**, 328–334 (1989).
- Hastie, N. *et al. Nature* **346**, 866–868 (1990).
- Harley, C. B., Futcher, C. W. & Greider, C. W. *Nature* **345**, 458–461 (1990).
- Lane, D. P. *Nature* **358**, 15–16 (1992).
- Lynch, H. T. *et al. Cancer Genet. Cytogenet.* **53**, 143–160 (1991).
- Mullis, K. B. & Faloona, F. A. *Meth. Enzym.* **155**, 335–350 (1987).
- Shibata, D. *et al. Am. J. Path.* **141**, 1–5 (1992).
- Shibata, D., Schaeffer, J. Li, Z., Capella, G. & Perucho, M. *J. natn. Cancer Inst.* (in the press).
- Sugarbaker, P., Gunderson, L. & Wittes, R. In *Cancer, Principles and Practice of Oncology* (eds De Vita, V., Hellman, S. & Rosenberg, S.) 814–815 (Lippincott, Philadelphia, 1985).
- Peinado, M. A., Fernandez-Renart, M., Capella, G., Wilson, L. & Perucho, M. *Int. J. Oncol.* **2**, 123–134 (1993).
- Vogelstein, B. *et al. Science* **244**, 207–211 (1989).

ACKNOWLEDGEMENTS. We thank M. McClelland and J. Welsh for helpful and encouraging discussions and L. Wilson for technical help. This work was supported by grants from the National Cancer Institute to M.P. M.A.P. was a fellow of the Spanish Ministry of Education. We wish to dedicate this work to the authors of ref. 8, especially the first and last, for being exemplar examples of practitioners of good science in the past and present eras, respectively.

A tetrameric DNA structure with protonated cytosine·cytosine base pairs

Kalle Gehring*, Jean-Louis Leroy & Maurice Guéron

Groupe de Biophysique de l'École Polytechnique et de l'URA D1254 du CNRS, 91128 Palaiseau, France

OLIGOMERS containing tracts of cytidine form hemiprotonated base pairs at acid pH and have been considered to be double-stranded. We have solved the structure of the DNA oligomer 5'-d(TCCCCC) at acid pH and find that it is a four-stranded complex in which two base-paired parallel-stranded duplexes are intimately associated, with their base pairs fully intercalated. The relative orientation of the duplexes is antiparallel, so that each base pair is face-to-face with its neighbours. The NMR spectrum displays only six spin systems, showing that the structure is highly symmetrical on the NMR timescale; the four strands are equivalent. A model derived by energy minimization and constrained molecular dynamics shows excellent compatibility with the observed nuclear Overhauser effects (NOEs) particularly for the very unusual inter-residue sugar-sugar NOEs H1'-H1', H1'-H2'' and H1'-H4'. These NOEs are probably diagnostic for such tetrameric structures.

Cytosine-protonated cytosine (C·C⁺) base pairs (Fig. 1a) were first observed in crystals of acetyl cytosine¹ and later in solutions of polycytidylic RNA^{2,3} and DNA⁴. X-ray diffraction of polycytidylic RNA fibres⁵ suggested a parallel duplex, but the same diffraction pattern was later observed at pH 7 and interpreted as being a single-stranded form⁶. A parallel structure containing a C·C⁺ base pair has also been observed in single crystals of the complex of ribo(CpA) and proflavin^{7,8}. In an NMR study of d(A₅C₅) at acid pH, the imino proton spectrum was observed and assigned to hemiprotonated C·C⁺ base pairs in a parallel duplex⁹.

As part of an investigation of the acid form of deoxycytidine oligomers^{10,11}, we have now determined the structure and stoichiometry of the complex formed by the hexamer d(TC₅).

* Present address: Laboratoire de RMN, ICSN, CNRS, 91198 Gif-sur-Yvette, France.

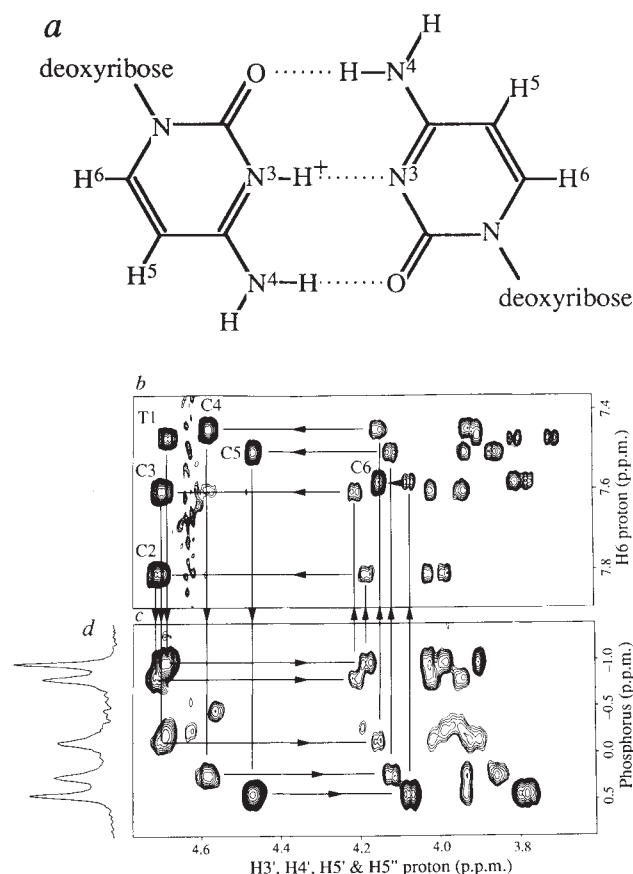


FIG. 1 Assignment of the phosphorus and sugar proton resonances of d(TC₅) in a heteronuclear total correlation spectroscopy (TOCSY) experiment. *a*, Scheme of hydrogen bonding in cytosine-protonated cytosine (C·C⁺) base pairs. The imino proton (on N3) switches rapidly between the two bases. *b*, Homonuclear NOESY experiment (240 ms) showing the six sugar proton spin systems (H3', H4', H5' and H5'' protons), each associated by NOE to its H6 base proton. All of the NOE crosspeaks are intra-residue. *c*, ¹H-³¹P hetero-TOCSY correlating the H3' and H5' and H5'' resonances with their directly coupled phosphorus resonances. (Starting from the 5' thymidine (H3', H6) crosspeak, labelled T1, follow the arrow down to the first (H3', ³¹P) crosspeak, across to the (H5', ³¹P) crosspeak, up to the (H5', H6) crosspeak of the following residue, and finally across to the (H3', H6) crosspeak, labelled C2.) The identification of several of the phosphorus resonances could be confirmed by (H4', ³¹P) and (H2', ³¹P) (not shown) TOCSY crosspeaks. A small amount of a possibly single-stranded contaminant is present as unassigned crosspeaks between -0.1 and -0.3 p.p.m. (³¹P dimension). *d*, One-dimensional ¹H-decoupled phosphorus spectrum of the tetrad d(TC₅) (162 MHz, 2,048 scans, 3 s recycle delay).

METHODS. 5'-d(TCCCC) (10 μmol) was synthesized by standard phosphoramidite chemistry on a Pharmacia GeneAssembler with a CPG resin (Millipore). After work-up and purification²⁶, the sample (200 optical density units as a Na⁺ salt) was dialysed against distilled H₂O and the pH lowered to 4.3 with HCl. Occasionally, slow cooling was necessary to eliminate unwanted conformers. Two-dimensional NMR spectra were acquired at 25 °C on an AMX600 spectrometer (Bruker Spectrospin). Strand concentration was 7.5 mM. The NOESY mixing time was 240 ms with ³¹P-decoupling in the *t*₂ dimension. The hetero-TOCSY used a DIPSI-2 mixing pulse sequence (6 loops, ~35 ms) with 32 TPPI increments, 800 scans per increment and a 2-s recycle delay¹⁴. Spectra were processed using FELIX 1.1 (Hare Research) with skewed sine-bell squared apodization functions, and plotted with a factor of 1.3 between contour levels.

As shown elsewhere, the spectrum of d(TC₅) in H₂O displays five imino proton resonances at 15 to 16 p.p.m., with an integrated area of one proton per pair of cytidines. The observation of these resonances is strong evidence of imino proton hydrogen-bonding. If it were accessible to water, the C⁺ imino proton would exchange in ~1 μs. In contrast, exchange broadening of most imino protons is negligible; indeed, the exchange times are as long as hours. The amino proton spectrum consists of only two lines, one from the five internal amino protons and one from the external protons. Based on this spectrum, it will be shown that each proton forms a hydrogen bond between two bases which are equivalent and must therefore be paired according to Fig. 1*a*. The lack of visible exchange broadening in the amino proton spectrum implies that the imino proton switches between the two bases at a rate >80,000 s⁻¹ (ref. 12).

The transition between the single-stranded and multistranded form can also be followed in the NMR spectrum of the non-exchangeable protons. The number of resonances in the single and multistranded forms is the same, which implies that all of the DNA strands in the multistranded form are identical on the NMR timescale. The sugar proton resonances of the multistranded acid form were assigned to one of the six nucleotide spin systems by COSY and TOCSY experiments using only through-bond *J*-couplings. These groupings are also found in the intra-residue NOEs (Fig. 1*b*), between the H6 base protons and the H3', H4', H5'/5'' protons. The H6 were assigned through these NOEs. The deoxyribose spin systems were aligned independently of any NOESY experiments¹³ by a heteronuclear ¹H-³¹P TOCSY experiment¹⁴. In the spectrum, each phosphorus resonance shows crosspeaks with the H3' resonance of the preceding nucleotide and with its own H5' and H5'' resonances (Fig. 1*c*).

NOESY experiments at mixing times of 30 and 240 ms showed very unusual inter-residue sugar-sugar NOEs (Fig. 2*A, B*; top). The assignment of the deoxyribose protons establishes that they are essentially all long-range NOEs (between residues distant

in the primary sequence). Similar long-range NOEs are found in many regions of the NOESY spectra: imino-imino, imino-H6, H6-H3', H1'-H2'', H1'-H4', H5-H1' and H6-H1'. In this last region, the sequence of NOEs can be read in its entirety as: T1-C6-C2-C5-C3-C4 (Fig. 2*A, B*, bottom). We term such NOEs 'pseudo-sequential' because they reflect the order of base-pair stacking in the DNA complex rather than the covalent sequence. To account for these NOEs, we initially considered a number of alternative DNA structures with different strand arrangements. By various arguments, it is possible to show that only the four-stranded model *k* is consistent with both the C·C⁺ base-pairing scheme and the observed equivalence of all the DNA strands (Fig. 2*C*).

Independent evidence for a tetrameric complex comes from dilution experiments using polyacrylamide gel electrophoresis. At pH 4.5, the oligonucleotide d(TC₅) migrates in a non-denaturing gel as three species whose relative proportions are strongly concentration-dependent (Fig. 3*a*). At concentrations above 100 μM, a slowly migrating species (T) dominates; at intermediate concentrations, a more rapidly migrating species (D) appears, and at concentrations below 1 μM the most rapidly migrating species (M) dominates. At thermodynamic equilibrium (achieved by slow cooling before electrophoresis), the stoichiometry of the T species can be measured by plotting the logarithm of its concentration against that of the presumed monomer species (M). From the slope of a line of best fit, we conclude that the slowly migrating T species is a complex of four strands (Fig. 3*b*).

A molecular model based on the strand arrangement of model *k* (Fig. 2*C*) was constructed using the molecular dynamics package, X-PLOR¹⁵. Inclusion of 70 NOE-distance and 10 torsion-angle constraints resulted in a family of closely related structures, of which a representative member is shown in Fig. 4. A more complete account of the structure determination will be published elsewhere.

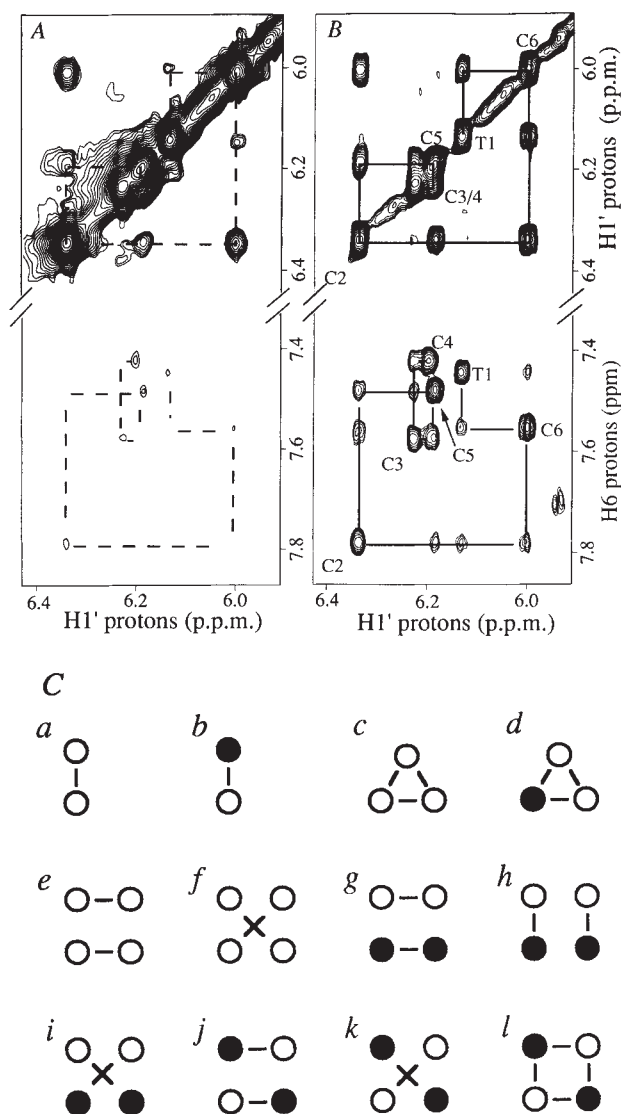
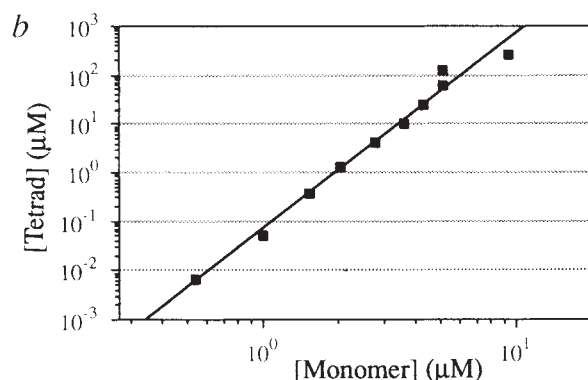
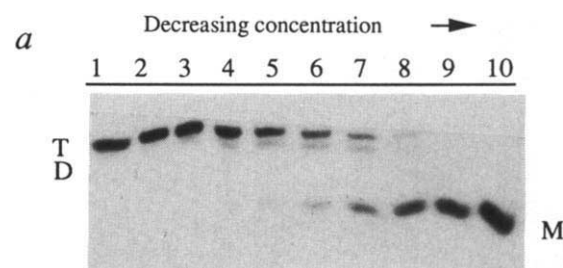


FIG. 2 Determination of the order of base-pair stacking in the complex from NOESY spectra at short (A, 30 ms) and long mixing times (B, 240 ms). Top portion of figures, the stacking order according to H1'-H1' pseudo-sequential NOE crosspeaks: T1-C6-C2-C5-C3/4. The 30-ms spectrum emphasizes direct cross-peaks between protons at close distance. Lower portion, H1'-H6 NOEs. Lines connect the intra-residue (labelled by residue) and inter-residue H6-H1' pseudo-sequential NOEs to give the complete order of base-pair stacking: T1-C6-C2-C5-C3-C4. The equal intensity of the H6-H1' (5' to 3') and H1'-H6 crosspeaks is a consequence of the face-to-face orientation of the base pairs. There is one true sequential crosspeak between T1 H1' and C2 H6. Spectra were acquired as for Fig. 1. C, Hypothetical strand arrangements for the complex. End-on views of dimeric (a, b), trimeric (c, d, e) and tetrameric structures (f-l) are shown with the polarity of the strands indicated by filled or open circles. Base pairing between strands is indicated by short lines. Not all possible structures are shown. Models with all strands parallel (a, c, e, f) are excluded by the long-range NOEs. Symmetry constraints, based on the NMR equivalence of all the DNA strands, eliminate alternative models: d, g, h and i (by consideration of each strand in stereo-specific relation to its neighbours). All the remaining models, except k, contain base pairs between antiparallel strands. Antiparallel base pairing with a symmetric (C-C⁺) base pair requires that one of the glycosidic bonds be *syn* and the other *anti*²⁷. This is excluded by the intense H6-H3' NOEs at a mixing time of 30 ms (data not shown) which show that all of the bases are in the *anti* orientation²⁸. These models (b, j, l) also fail to account for the number of imino resonances. All antiparallel models necessarily have an axis of 2-fold symmetry between the central base pairs (because the strands are identical). As the imino protons are in rapid exchange between the bases in a C-C⁺ base pair, this would halve the number of imino proton resonances. The only remaining strand arrangement is model k.

FIG. 3 Concentration dependence of tetrad formation. a, Autoradiogram of twofold dilutions of d(TC₅) analysed by gel electrophoresis. Total oligomer concentrations (in strands) ranged from 280 μ M (lane 1) to 0.56 μ M (lane 10). T, d(TC₅) tetrad; D, a presumed duplex form; M, monomer (see text). b, Double logarithmic plot of the concentrations (in strands) of the M and T species. The line of least-squares linear regression has a slope of 4.0, with a tetrad dissociation constant of 55 (μ M)³ ($\Delta G^\circ = -21.2 \pm 0.1$ kcal mol⁻¹). The estimated ΔG° for duplex formation is -6.5 ± 0.5 kcal mol⁻¹. METHODS. Serial twofold dilutions of d(TC₅) were mixed with a small amount of 5'-³²P-labelled d(TC₅) and allowed to equilibrate in (80 mM Tris-acetate, 1 mM EDTA, pH 4.5) at 10 °C before gel electrophoresis on a 20% non-denaturing polyacrylamide gel (run at 2 °C for 6 h at 10 V per cm). The relative proportions of radioactive decay in the T, D and M species in each lane were measured on a Molecular Dynamics PhosphorImager and the concentrations of each species calculated. Depending on the methods used for base-line correction and integration, the calculated stoichiometry (least-squares slope) varied between 3.5 and 4.0 for the T species alone, and between 3.2 and 3.5 when the D and T species were pooled. Gel electrophoresis at pH 8 shows only a single species at all concentrations (results not shown).



The structure consists of two parallel duplexes (antiparallel to one another) with twelve intercalating base pairs (Fig. 4a). The dyad symmetry along the tetrad axis (seen on the NMR timescale) limits the possible base positions so that the centres of the ten C·C⁺ base pairs are collinear, with successive base pairs offset by a rise and rotation about the tetrad axis. Shifts, rolls and tilts of the base pairs are excluded; only propeller twist and base-pair buckling¹⁶ are allowed. Figure 4d shows a detail of two stacked base pairs, C3·C3 and C4·C4, viewed along the tetrad axis. They belong to distinct duplexes and are perpendicular to one another and face-to-face. The sugar-phosphate backbones of the tetrad are associated pairwise with close van der Waals contacts between the sugars (Fig. 4c). The H1' protons are essentially above one another, separated on average by 0.30 nm. In the couples, C4·C4, C3·C5 and C2·C6, the inter-residue H1'–H4' and H4'–H4' distances are short, whereas in the alternate couples, C4·C3, C5·C2, C6·T1, the H1'–H2' and H2'–H2' distances are short.

The tetrad helical twist is small and right-handed, with about

16° between base pairs belonging to the same duplex. The phosphate groups are well separated (by more than 0.6 nm), except for phosphates C4 and C5, which are separated by only 0.49 nm (Fig. 4c). The backbone dihedral angles fall into two classes. For residues C2 and C3, the torsion angle γ is *gauche*⁺, as in B-form DNA, and the pseudo-rotation angle P is N-type. This is reflected¹⁷ in the absence of H4'–H5'/5" and H1'–H2' scalar couplings (data not shown). For residues C4, C5 and C6, γ is *trans* and the pseudo-rotation angle P is shifted towards C4' *exo*. This diversity in P is visible in the line shapes of the H1' and H3' resonances (Figs 2B and 1b). The other backbone torsion angles are presumably correlated with γ and P , and their variation may contribute to the chemical shift dispersion in the phosphorus spectrum (Fig. 1d).

Compared with a duplex, the tetrad presents several energetic advantages. First, the tetrad has 11-base-pair stacking interactions, compared to only 10 for two duplexes. Second, the C·C⁺ pairs carry a positive charge on the C⁺ imino proton and ring nitrogen N3 (Fig. 1a). In a duplex, this would favour proton-

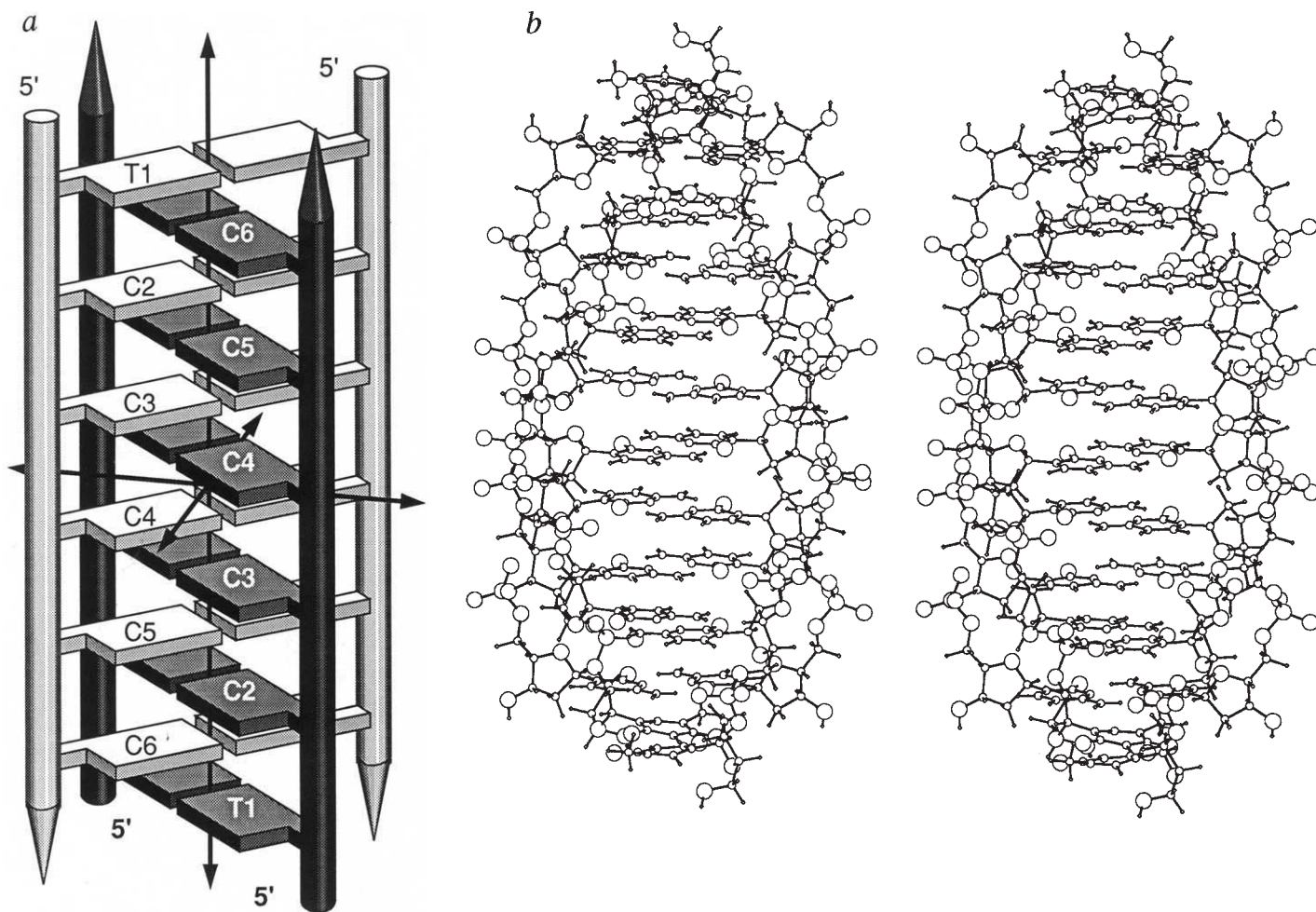


FIG. 4 Proposed model of the d(TC₅) tetrad. *a*, Intercalation scheme of the two parallel duplexes. The terminal thymidine residues may not be base-paired. There are three axes of 2-fold symmetry (arrows); one along the helix axis through the centre of each C·C⁺ base pair, and two between the central C4·C4 base pairs. *b*, Stereo view of an energy-minimized structure calculated using 70 NOE distance and 10 dihedral angle constraints. *c*, Side view of the structure showing two of the sugar-phosphate backbones and the many close van der Waals contacts between sugars. These give rise to the strong inter-residue H1'–H1', H1'–H2" and H1'–H4' NOEs. *d*, Stacked C·C⁺ base pairs, C4·C4 (red, in front) and C3·C3 (blue, behind). The C⁺ imino protons (in fast exchange between the bases) are not shown. The 5' phosphates of C4·C4 are above the bases and the C3·C3 phosphates below. Protons C3 H1' and C4 H2" are separated by only 0.22 nm.

METHODS. NOE crosspeak volumes (38 intra-residue, 5 true sequential, 7 long-range) at 30 ms were converted to distances using cytosine H5–H6 and deoxyribose H2'–H2" and H5'–H5" distances as references. At 240 ms, additional NOE crosspeaks (5 intra-residue, 1 sequential, 14 long-range) were estimated by eye to be strong, medium or weak, and converted to distances of less than 0.4, 0.6 and 0.7 nm, respectively. Dihedral angle constraints were added for the γ -backbone angles of residues C2 (*gauche*⁺), C3 (*gauche*⁺), C5 (*trans*), and C6 (*trans*) and for the six sugar-pucker pseudo-rotation angles P . The input structure was a four-stranded complex built with B-form DNA strands and the base-pairing scheme in *a*. Hydrogen bonds between cytosines were included as NOE and dihedral angle constraints. In the final structure, there were no NOE violations of over 0.01 nm.

ation on alternate bases in order to maximize the distance between charged cytosines. Movement of the imino proton from one strand to the other would be greatly slowed, as in the crystal structure of 1-methylcytosine¹⁸. (However, symmetric C·C⁺ pairs have been observed^{1,7}). In the tetrad, the cytosine N3 nitrogens are maximally separated, allowing disorder in the pattern of protonation with a concomitant free energy gain. A further electrostatic advantage comes from the reversed orientation of the carbonyl and amine dipoles in successive base pairs (Fig. 4d). Last, the tetrad presents extensive sugar-sugar van der Waals contacts that would be absent in a C·C⁺ duplex.

The structure reported here includes a new motif for nucleic acids: intercalated base pairs. This motif is designated the 'i-motif'. It seems to form readily in sequences containing stretches of deoxycytidine, because it has now been partly characterized in several other sequences. For instance, d(T₂C₈T₂), d(TC₃), d(TC₃T) and d(C₄TC₄) form tetramers (as shown by gel filtration chromatography¹² and they display the unusual H1'-H1' NOESY crosspeaks (not shown). The i-motif could possibly be a common feature of structures containing parallel strands. In

contrast, the merging of two antiparallel duplexes of Watson-Crick pairs would necessitate placing one sugar-phosphate backbone of one duplex within the minor groove of the other, which seems impossible.

In our conditions, the d(TC₃) tetrad is not stable at pH 7.0. But, in other conditions, this or other oligomers containing cytidine sequences might produce tetrads at physiological pH and be biologically important. In particular, the C-rich strand of telomeres, whose structure is of current interest^{19,20}, includes extensively repeated sequences such as C₃A₂, C₄A₂ or C₄A₄ (ref. 21). A length of single strand consisting of four consecutive sequences such as these could fold to form the i-motif, analogous to the formation of G-tetrads by sequences representative of the other strand²²⁻²⁵. □

Received 15 December 1992; accepted 23 March 1993.

1. Marsh, R. E., Bierstedt, R. & Eichhorn, E. L. *Acta crystallogr.* **15**, 310-316 (1962).
2. Akrimimiski, E. O., Sander, C. & Ts'o, P. O. P. *Biochemistry* **2**, 340-344 (1963).
3. Hartman, K. A. Jr & Rich, A. *J. Am. Chem. Soc.* **87**, 2033-2039 (1965).
4. Inman, R. B. *J. molec. Biol.* **9**, 624-637 (1964).
5. Langridge, R. & Rich, A. *Nature* **198**, 725-728 (1963).
6. Arnott, S., Chandrasekaran, R. & Leslie, A. G. W. *J. molec. Biol.* **106**, 735-748 (1976).
7. Westhof, E., Roa, S. T. & Sundaralingam, M. *J. molec. Biol.* **142**, 331-361 (1980).
8. Westhof, E. & Sundaralingam, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1852-1856 (1980).
9. Luo, J., Sarma, M. H., Yuan, R. & Sarma, R. H. *FEBS Lett.* **306**, 223-228 (1992).
10. Leroy, J.-L., Kettani, A., Kochoyan, M. & Guéron, M. *Abstr. XIV Int. Conf. Magn. Reson. in Biol. Systems, Warwick* (1990).
11. Leroy, J.-L., Gehring, K. & Guéron, M. *Abstr. XV Int. Conf. Magn. Reson. in Biol. Systems, Jerusalem* (1992).
12. Leroy, J.-L. *et al. Biochemistry* (in the press).
13. Aboul-ela, F., Murchie, A. I. H. & Lilley, D. M. J. *Nature* **360**, 280-282 (1992).
14. Kellogg, G. W. *J. magn. Reson.* **98**, 176-182 (1992).
15. Brünger, A. T. *X-PLOR (Version 2.1) Manual* (Yale Univ., Connecticut, 1990).
16. EMBO Workshop, *EMBO J.* **8**, 1-4 (1989).
17. Altona, C. *Recl. Chim. Pays Bas* **101**, 413-433 (1982).
18. Kistenmacher, T. J., Rossi, M. & Marzilli, L. G. *Biopolymers* **17**, 2581-2585 (1978).
19. Lyamichev, V. I. *et al. Nature* **339**, 634-637 (1989).
20. Ahmed, S. & Henderson, E. *Nucleic Acids Res.* **20**, 507-511 (1992).
21. Blackburn, E. H. *Nature* **350**, 569-572 (1991).
22. Sen, D. & Gilbert, W. *Nature* **334**, 364-366 (1988).
23. Sundquist, W. I. & Klug, A. *Nature* **342**, 825-829 (1989).
24. Smith, F. W. & Feigon, J. *Nature* **356**, 164-168 (1992).
25. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. *Nature* **356**, 126-131 (1992).
26. Kochoyan, M., Leroy, J.-L. & Guéron, M. *Biochemistry* **29**, 4799-4805 (1990).
27. Westhof, E. *Nature* **358**, 459 (1992).
28. Wüthrich, K. *NMR of Proteins and Nucleic Acids* 211 (Wiley, New York, 1986).

ACKNOWLEDGEMENTS. We thank A. Morillon for oligomer synthesis; C. Reisdorf and the Département de Chimie, École Polytechnique, for assistance and for spectrometer time; M. Kochoyan for help with the modelling; N. Varin-Blank of the Pasteur Institute for help with the Molecular Dynamics Phosphorimager; and Y. Mechulam and other members of the Laboratoire de Biochimie, École Polytechnique for help with gel electrophoresis. Parts of this work were carried out on the 360 Mhz spectrometer of the Groupe de Biophysique and processed with an early version of FELIX provided by D. Hare of Hare Research Inc. This study was supported by a grant from the Association pour la Recherche contre le Cancer (M.G.); partial support (K.G.) was provided by the École Polytechnique and by the French Ministère des Affaires Étrangères.

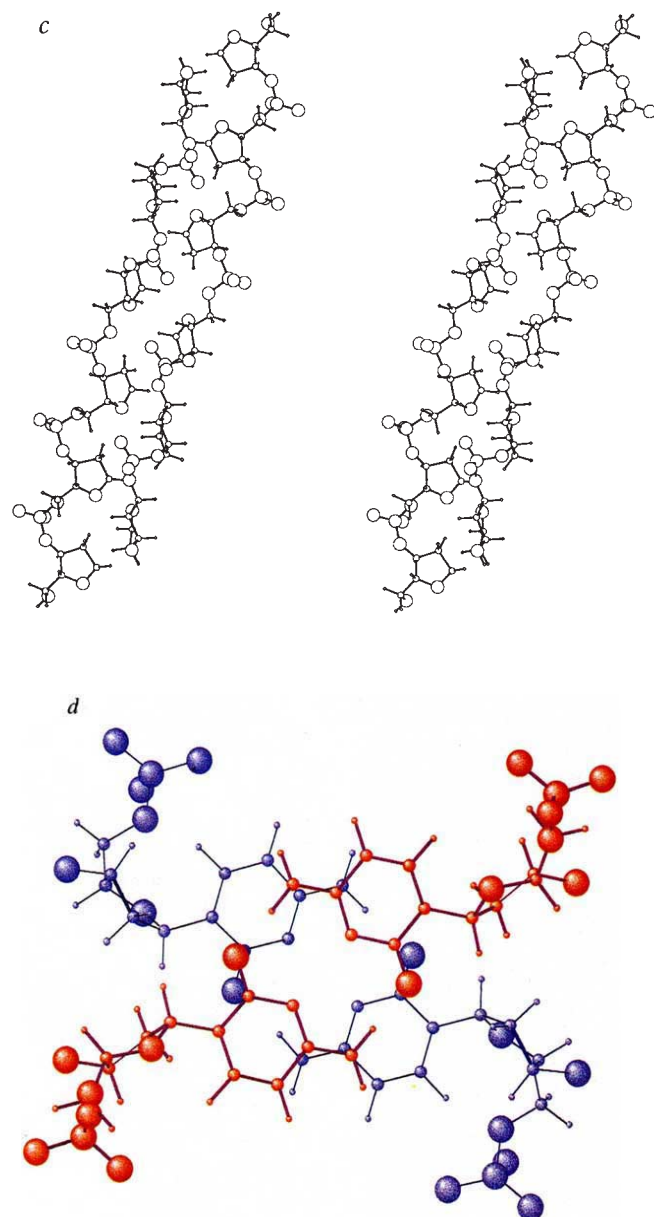
ERRATUM

Microsecond-resolved XAFS of the triplet excited state of Pt₂(P₂O₅H₂)₄⁴⁻

D. J. Thiel, P. Livins, E. A. Stern & A. Lewis

Nature **362**, 40-43 (1993).

THE penultimate sentence of the first introductory paragraph of this letter should read, "We find that the triplet excited state of Pt₂(P₂O₅H₂)₄⁴⁻, with a lifetime of about 4 μs, undergoes a contraction in the separation of the P-planes of 0.52 ± 0.13 Å relative to the ground state." As published, the contraction was wrongly attributed to the Pt-Pt distance. □



PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to presubmission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and should contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the Commentary Editor in the form of a one-page synopsis.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, and clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.